

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 21:00:09 ; Search time 1434 Seconds

(without alignments)
7322.251 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gaattgcgccccttctaatacg.....aaaaaaaaaagcgccgc 1859

Scoring table: ORIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4176236 seqs, 2824127955 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1859	100.0	1859	10	US-09-993-966-5 Sequence 5, Appl1
2	1786	96.1	1786	10	US-09-993-966-1 Sequence 1, Appl1
3	1417	76.2	1448	13	US-10-087-192-179 Sequence 179, App
4	783	42.1	2379	10	US-09-764-891-5902 Sequence 5902, Ap
5	783	42.1	2379	10	US-09-764-891-5904 Sequence 5904, Ap
6	783	42.1	106315	13	US-10-087-192-178 Sequence 178, App
7	590	31.7	590	10	US-09-993-966-21 Sequence 21, Appl
8	377	20.3	427	10	US-09-918-995-8686 Sequence 8686, Ap
9	215	11.6	215	10	US-09-993-966-22 Sequence 22, Appl
10	148	8.0	148	10	US-09-993-966-18 Sequence 18, Appl
11	137	7.4	598	15	US-10-029-386-7990 Sequence 7990, Ap
12	137	7.4	1950	13	US-10-016-157A-126 Sequence 126, App

13	134	7.2	134	10	US-09-993-966-14 Sequence 14, Appl
14	133	7.2	133	15	US-10-029-386-21690 Sequence 21690, A
15	128	6.9	128	10	US-09-993-966-20 Sequence 20, Appl
16	120	6.5	120	10	US-09-993-966-10 Sequence 10, Appl
17	108	5.8	639	15	US-10-243-552-231 Sequence 231, App
18	107	5.8	107	10	US-09-993-966-16 Sequence 16, Appl
19	96	5.2	96	10	US-09-993-966-17 Sequence 17, Appl
20	88	4.7	554	15	US-10-029-386-7615 Sequence 7615, Ap
21	86	4.6	179	15	US-10-029-386-21315 Sequence 21315, A
22	85	4.6	85	10	US-09-993-966-19 Sequence 19, Appl
23	68	3.7	566	13	US-10-027-632-266931 Sequence 266931, Sequence 266931,
24	68	3.7	566	15	US-10-027-632-266931 Sequence 1214, Ap
25	68	3.7	2724	9	US-09-833-381-1215 Sequence 1215, Ap
26	68	3.7	2724	9	US-09-833-381-1215 Sequence 15, Appl
27	67	3.6	67	10	US-09-993-966-15 Sequence 15, Appl
28	58	3.1	670	16	US-10-275-287-19 Sequence 19, Appl
29	57	3.1	417	16	US-10-275-287-17 Sequence 17, Appl
30	57	3.1	510	16	US-10-275-287-17 Sequence 25, Appl
31	56	3.0	636	16	US-10-275-287-25 Sequence 25, Appl
32	56	3.0	651	16	US-10-275-287-8 Sequence 8, Appl1
33	56	3.0	686	16	US-10-275-287-18 Sequence 18, Appl
34	56	3.0	707	15	US-10-072-602B-214 Sequence 214, App
35	55	3.0	547	16	US-10-275-287-10 Sequence 10, Appl
36	55	3.0	1245	16	US-10-275-287-52 Sequence 52, Appl
37	55	3.0	1245	16	US-10-275-287-16 Sequence 16, Appl
38	54	2.9	1874	17	US-10-635-223-16 Sequence 20, Appl
39	54	2.9	1874	17	US-10-635-223-42 Sequence 42, Appl
40	54	2.9	2605	17	US-10-635-223-42 Sequence 54, Appl
41	53	2.9	787	16	US-10-275-287-54 Sequence 27, Appl
42	53	2.9	1434	15	US-10-210-760A-27 Sequence 122, App
43	53	2.9	1434	16	US-10-229-541A-122 Sequence 13, Appl
44	53	2.9	3281	15	US-10-171-319-13 Sequence 949, App
45	53	2.9	3768	14	US-10-101-464A-949

ALIGNMENTS

```

RESULT 1
US-09-993-966-5
; Sequence 5, Application US/09993966
; Publication No. US20030186232A1
GENERAL INFORMATION:
APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,
; FILE REFERENCE: 014024/0280733
CURRENT APPLICATION NUMBER: US/09/993, 966
CURRENT FILING DATE: 2001-11-27
PRIOR APPLICATION NUMBER: 60/252, 884
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/291, 109
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/325, 571
PRIOR FILING DATE: 2001-10-01
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-5
Query Match 100.0%; Score 1859; DB 10; Length 1859;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCGCCCTTCTAATACACTCTATPAGGCAAGAGTGTAAACAAGCAGAGTACG 60
Db 1 GAATTCGCCCTTCTAATACACTCTATPAGGCAAGAGTGTAAACAAGCAGAGTACG 60
QY 61 CGGGAGTGGCGCGCGGCGGAGCGGCGGAGCGGCGGTCCCGGCGGCGGTCCG 120

```

Db	6	CGGGAGATCGGGGCGCGCGCGACCGCGCGAGAGCGCGTCCGCGCGCGCGCTTCGGGACTCCG	120
Qy	121	CTCGGCTTCGGGGGCTGCTTCGGGAGAGAGAGAGACCCAAAGAGAGGGCCGAGGGCCCGGGGCC	180
Db	121	CTCGGCTTCGGGGGCTGCTTCGGGAGAGAGAGAGACCCAAAGAGAGGGCGCCAGGCTCCGGGCC	180
Qy	181	GAGGCGCATATGCTTATGAGGACGCTCCCGGGCCGCGAGGCCAGGCTCGGAGAACTTCACTC	240
Db	181	GAGGCGCATATGCTTATGAGGACGCTCCCGGGCCGCGAGGCCAGGCTCGGAGAACTTCACTC	240
Qy	241	CAAGCCGAGCGCCGCTGTGTCAAAGCCGAGGAGAGCCCGGAGTGAATCAAGCTTTCGCCGTGAG	300
Db	241	CAAGCCGAGCGCCGCTGTGTCAAAGCCGAGGAGAGCCCGGAGTGAATCAAGCTTTCGCCGTGAG	300
Qy	301	CGCTGCGCTGGGCTCGGAAAGGCGCATTCGAGGATGTGATTCGGGAGACAGGCGCTGCCGGGCGG	360
Db	301	CGCTGCGCTGGGCTCGGAAAGGCGCATTCGAGGATGTGATTCGGGAGACAGGCGCTGCCGGGCGG	360
Qy	361	TGTCCTCGGAGACCCCGACAGCTTCGGATTGGCGGGGACCATAGGCCGAAGCACCCGGAGACT	420
Db	361	TGTCCTCGGAGACCCCGACAGCTTCGGATTGGCGGGGACCATAGGCCGAAGCACCCGGAGACT	420
Qy	421	CGTGGGCGACGTGTATGAGAGACAGGCTTCAGGAGAGAAAGAGAGACGACTTCGGCTGCGA	480
Db	421	CGTGGGCGACGTGTATGAGAGACAGGCTTCAGGAGAGAAAGAGAGACGACTTCGGCTGCGA	480
Qy	481	AGTGGCCCTGCGCTCCTTGAGAGACTGACGCGCTGGGCGAGCGAGATGAGAGAAAGATGGA	540
Db	481	AGTGGCCCTGCGCTCCTTGAGAGACTGACGCGCTGGGCGAGCGAGATGAGAGAAAGATGGA	540
Qy	541	GAGAGTAGAGCAACCTTCGCCAGGCTCCAAAGACGCTGAAGTTTGAAGGCTTCAGTG	600
Db	541	GAGAGTAGAGCAACCTTCGCCAGGCTCCAAAGACGCTGAAGTTTGAAGGCTTCAGTG	600
Qy	601	CGAGGTGTCATATGAGAGAGAGACAGCGCGGCAAGATGAGACTTCAACCTGTATGACTTTGA	660
Db	601	CGAGGTGTCATATGAGAGAGAGACAGCGCGGCAAGATGAGACTTCAACCTGTATGACTTTGA	660
Qy	661	CAACAAACGGCAAGGTCACCCGAGAGAGACATCAACGACTTGTGTCACACCATCTATAGAGT	720
Db	661	CAACAAACGGCAAGGTCACCCGAGAGAGACATCAACGACTTGTGTCACACCATCTATAGAGT	720
Qy	721	GGTGGACTCCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTGCGGGTAAAGCTTAC	780
Db	721	GGTGGACTCCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTGCGGGTAAAGCTTAC	780
Qy	781	CGTGGCCCCCGATGGCAGCCAGAGCAAGAGGCGTCTTGTCAATCAAGCTGACCTTGA	840
Db	781	CGTGGCCCCCGATGGCAGCCAGAGCAAGAGGCGTCTTGTCAATCAAGCTGACCTTGA	840
Qy	841	GAGGCGCAAGGCCCCGAGAGAGAGACCAAGCCCATGTAGAGACTTGGCGGAGCTGGAGAGAA	900
Db	841	GAGGCGCAAGGCCCCGAGAGAGAGACCAAGCCCATGTAGAGACTTGGCGGAGCTGGAGAGAA	900
Qy	901	GCAGCGAGCCCGGCTCAGGTTTCAGAGGTGACAGCGCTGAGAGCAAGTGTGGCTGTACCA	960
Db	901	GCAGCGAGCCCGGCTCAGGTTTCAGAGGTGACAGCGCTGAGAGCAAGTGTGGCTGTACCA	960
Qy	961	CCATTGCGTATGAGAGCAATCGAGAGAGAAACCACTACTTAGATCTTCGCCGGGATGGA	1020
Db	961	CCATTGCGTATGAGAGCAATCGAGAGAGAAACCACTACTTAGATCTTCGCCGGGATGGA	1020
Qy	1021	AAACTGACAGTCCCAATTTGGGCGCTGCTCCCGTTCCGTGGCCCGAGAGTCAAGAACTGCG	1080
Db	1021	AAACTGACAGTCCCAATTTGGGCGCTGCTCCCGTTCCGTGGCCCGAGAGTCAAGAACTGCG	1080
Qy	1081	CCCCCGACCTTCAATTCACACTCTCGCTCCCATAGAGCCGGAAGCATTCACATCC	1140
Db	1081	CCCCCGACCTTCAATTCACACTCTCGCTCCCATAGAGCCGGAAGCATTCACATCC	1140
Qy	1141	ACACCGAAAGCCCAAGGCGTGAACCCGAGCTCTTTCACATTCCTTGAACCCCAATGCG	1200
Db	1141	ACACCGAAAGCCCAAGGCGTGAACCCGAGCTCTTTCACATTCCTTGAACCCCAATGCG	1200

Qy	1201	CAAGGTCACAGACTCCAGGAACGCTCCGGGGGACCCAGAGAGGGGAGCAAGCACTTGT	1260
Db	1201	CAAGGTCACAGACTCCAGGAACGCGCTCCGGGGACCCAGAGAGGGGAGCAAGCACTTGT	1260
Qy	1261	GAGGTCCCCCAAGAGCCACGAGGGCAAGAGTGTGTGGGACACGTGACACAGAGGGCAAG	1320
Db	1261	GAGGTCCCCCAAGAGCCACGAGGGCAAGAGTGTGTGGGACACGTGACACAGAGGGCAAG	1320
Qy	1321	AAACAMGCCCCCTCTGTGGGACCCGACATCCCTGGGGTGTCCCCCTCCGGCCACCTGGCTGC	1380
Db	1321	AAACAMGCCCCCTCTGTGGGACCCGACATCCCTGGGGTGTCCCCCTCCGGCCACCTGGCTGC	1380
Qy	1381	CAGCCGGGCGCTCTCCCTCCCTAGGCCCCCTCGGGGCAAAAGAGAGCAAGACACGAGC	1440
Db	1381	CAGCCGGGCGCTCTCTCCCTCTAGGCCCCCTCGGGGCAAAAGAGAGCAAGACACGAGC	1440
Qy	1441	CAAGAGAGGCGACAGAGGGCTGCCGGGGCTTGCAGAGCCACCTGGCTCAGGTGGCCTGT	1500
Db	1441	CAAGAGAGGCGACAGAGGGCTGCCGGGGCTTGCAGAGCCACCTGGCTCAGGTGGCCTGT	1500
Qy	1501	CTTGGGGCGGGAGACCTGGGGGAGCTGCCCGCTTGTTGGTGTATGAGGACGAGGCGG	1560
Db	1501	CTTGGGGCGGGAGACCTGGGGGAGCTGCCCGCTTGTTGGTGTATGAGGACGAGGCGG	1560
Qy	1561	GCAGCCGGTCACAGACATGAGACACCAACCAACATCACCCACATTACCAACA	1620
Db	1561	GCAGCCGGTCACAGAGATGAGACACCAACCAACATGAAATCACCCATTACCAACA	1620
Qy	1621	CTTCTACCAAGACATGAGCCCTCTCCAGGCCCCCACCTGCCCTATGAAAGAACCCAC	1680
Db	1621	CTTCTACCAAGACATGAGCCCTCTCCAGGCCCCCACCTGCCCTATGAAAGAACCCAC	1680
Qy	1681	CCGAGACCAACAAGGCACTTATTTATTCATTAATATATGTTATTAATGATATTTAT	1740
Db	1681	CCGAGACCAACAAGGCACTTATTTATTCATTAATATATGTTATTTATGATATTTAT	1740
Qy	1741	TAATATTAATTTGTTACTCCACTAATATTTAGCTAGCTACATGTAGAAAGATCTATGAAA	1800
Db	1741	TAATATTAATTTGTTACTCCACTAATATTTAGCTAGCTACATGTAGAAAGATCTATGAAA	1800
Qy	1801	CACGAACTAAACTTTTATTTATTTATGTTAAAAAAAAAAAAAAAAAAAAGCGGCGCG	1859
Db	1801	CACGAACTAAACTTTTATTTATTTATGTTAAAAAAAAAAAAAAAAAAAAGCGGCGCG	1859

```

RESULT 2
US-09-993-966-1
; Sequence 1, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NMD PROTEIN
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-1

Query Match      96.1%; Score 1786; DB 10; Length 1786;
Best Local Similarity 100.0%; Pctd. No. 0;

```

Matches 1786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	66	AGTCGGGCGCGGCGAGCGGCGAGAGCGGCTCCCGCGCGCTTGCGGCTCCGCTCGG	125	
Db	1	AGTCGGGCGCGGCGAGCGGCGAGAGCGGCTCCCGCGCGCTTGCGGCTCCGCTCGG	60	
Qy	126	CTGGGGGGGCTGCTTCGGGAGGAGGAGGCGCAAGGAGGAGGCGCGAGCGCGGGGGGGGG	185	
Db	61	CTGGGGGGGCTGCTTCGGGAGGAGGAGGCGCAAGGAGGAGGCGCGAGCGCGGGGGGG	120	
Qy	186	CATGGCTTAAGGAGCGCTCCCGCGCGCGCGAGCGCGCGAGCGCGAGCGCGAGCGCG	245	
Db	121	CATGGCTTAAGGAGCGCTCCCGCGCGCGCGAGCGCGCGAGCGCGAGCGCGAGCGCG	180	
Qy	246	CGGCGCGCGCTGCTGAGCGGAGGAGGAGCGCGAGGAGGAGCGCGAGGAGGAGCGCG	305	
Db	181	CGGCGCGCGCTGCTGAGCGGAGGAGGAGCGCGAGGAGGAGCGCGAGGAGGAGCGCG	240	
Qy	306	CTGGGGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	365	
Db	241	CTGGGGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	300	
Qy	366	CGGAGCGCGGAGCGCTCGGCTCGGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG	425	
Db	301	CGGAGCGCGGAGCGCTCGGCTCGGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG	360	
Qy	426	GCGAGCTGTTGAGAGACGCTCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	485	
Db	361	GCGAGCTGTTGAGAGACGCTCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	420	
Qy	486	CGCTGCTCTCGGAGAGACGCGCGGCTGGGCGAGCGCGAGCGCGAGCGCGAGCGCG	545	
Db	421	CGCTGCTCTCGGAGAGACGCGCGGCTGGGCGAGCGCGAGCGCGAGCGCGAGCGCG	480	
Qy	546	TGAGCGGAGCGCTCGGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG	605	
Db	481	TGAGCGGAGCGCTCGGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG	540	
Qy	606	TGTCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	665	
Db	541	TGTCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	600	
Qy	666	ACGGGAGGAGTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	725	
Db	601	ACGGGAGGAGTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	660	
Qy	726	ACTGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	785	
Db	661	ACTGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	720	
Qy	786	CGCCCGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	845	
Db	721	CGCCCGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	780	
Qy	846	CAAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	905	
Db	781	CAAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	840	
Qy	906	GAGCGCGGCTCAGGTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	965	
Db	841	GAGCGCGGCTCAGGTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	900	
Qy	966	GCGTAGATGAGAACATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1025	
Db	901	GCGTAGATGAGAACATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	960	
Qy	1026	ACAAGTCCCAATTTGGGCGCTGCTCCCTTCCTCGGCGCGAGGAGGAGGAGGAGGAGG	1085	
Db	961	ACAAGTCCCAATTTGGGCGCTGCTCCCTTCCTCGGCGCGAGGAGGAGGAGGAGGAGG	1020	
Qy	1086	GCAAGTCCCAATTTGGGCGCTGCTCCCTTCCTCGGCGCGAGGAGGAGGAGGAGGAGG	1145	
Db	1021	GCAAGTCCCAATTTGGGCGCTGCTCCCTTCCTCGGCGCGAGGAGGAGGAGGAGGAGG	1080	

Qy	1146	GAAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1205	
Db	1081	GAAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1140	
Qy	1206	TCTCAGAGCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1265	
Db	1141	TCTCAGAGCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1200	
Qy	1266	CGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1325	
Db	1201	CGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1260	
Qy	1326	AGGCGCGCTTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1385	
Db	1261	AGGCGCGCTTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1320	
Qy	1386	CGGCGCTCTCTCCCTCCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1445	
Db	1321	CGGCGCTCTCTCCCTCCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1380	
Qy	1446	AGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1505	
Db	1381	AGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1440	
Qy	1506	GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1565	
Db	1441	GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1500	
Qy	1566	CGGTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1625	
Db	1501	CGGTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1560	
Qy	1626	ACGAGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1685	
Db	1561	ACGAGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1620	
Qy	1686	CACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1745	
Db	1621	CACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1680	
Qy	1746	ATTATGTTACTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1805	
Db	1681	ATTATGTTACTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1740	
Qy	1806	AACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1851	
Db	1741	AACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1786	

RESULT 3
 US-10-087-192-179
 ; Sequence 179, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Engelhard, Eric K.
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
 ; FILE REFERENCE: 52945200122
 ; CURRENT APPLICATION NUMBER: US/10/087,192
 ; CURRENT FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 09/747,377
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: US 09/798,586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ. ID NOS: 2059
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ. ID NO 179
 ; LENGTH: 1448
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-087-192-179

Query Match 76.2%; Score 1417; DB 13; Length 1448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 CCCAGCATGGGAACTTCCTCAAGCCGCGCCGTGTGCAAGCGAGGAGAGCCC 276
DB 32 CCCAGCATGGGAACTTCCTCAAGCCGCGCCGTGTGCAAGCGAGGAGAGCCC 91
QY 277 GGAAGGTGACAGCTTCGCGGTGAGCGCTGCTGGGCTTCGAGAGGGCATTCGAGAGTGGAT 336
DB 92 GGAAGGTGACAGCTTCGCGGTGAGCGCTGCTGGGCTTCGAGAGGGCATTCGAGAGTGGAT 151
QY 337 CGGAGGACAGCGCTGCGCGCGGTGTCTCGGAGACCCCGACAGTGGGGTTGGGGGAC 396
DB 152 CGGAGGACAGCGCTGCGCGCGGTGTCTCGGAGACCCCGACAGTGGGGTTGGGGGAC 211
QY 397 CATAGCCGAAGCAACCCGGGAGCTCGTGGGCGAGCTGTGGAGAGACAGCTCAGCGAGGA 456
DB 212 CATAGCCGAAGCAACCCGGGAGCTCGTGGGCGAGCTGTGGAGAGACAGCTCAGCGAGGA 271
QY 457 AGAGAGAGACGACTTTCGGCTGGAAGTGGCCCTGCTTCGAGAAACTGAAGGGCTGGG 516
DB 272 AGAGAGAGACGACTTTCGGCTGGAAGTGGCCCTGCTTCGAGAAACTGAAGGGCTGGG 331
QY 517 CAGCGGAGATGAGAAGAAAGATGAGAGAGGGAACCCCGCAGGCTCCAGAGCA 576
DB 332 CAGCGGAGATGAGAAGAAAGATGAGAGAGGGAACCCCGCAGGCTCCAGAGCA 391
QY 577 GCTGAAGTTTGAAGAGCTCCAGTGCAGAGTGTCCATGAGAGAGAGACCCGCGAGAGTG 636
DB 392 GCTGAAGTTTGAAGAGCTCCAGTGCAGAGTGTCCATGAGAGAGAGACCCGCGAGAGTG 451
QY 637 GACCTTCAACCTGTATGACTTTTGAACAACGCGAGAGTCAACCGAGAGACATCACAG 696
DB 452 GACCTTCAACCTGTATGACTTTTGAACAACGCGAGAGTCAACCGAGAGACATCACAG 511
QY 697 CTGGCTGCAACCACTATGAGGTGTGTGACTCTCTGTCAACCACTCCCAATCCAG 756
DB 512 CTGGCTGCAACCACTATGAGGTGTGTGACTCTCTGTCAACCACTCCCAATCCAG 571
QY 757 CAAGATCTGCGGGTAAAGCTCAACCTGCGCCCGATGAGCGCAGCAGAGAGAGAGGT 816
DB 572 CAAGATCTGCGGGTAAAGCTCAACCTGCGCCCGATGAGCGCAGCAGAGAGAGAGGT 631
QY 817 CTTGTCAATCAGGCTGCTGCAAGCGCAAGGCCCGCAGCAGAGACCAAGCCACTGA 876
DB 632 CTTGTCAATCAGGCTGCTGCAAGCGCAAGGCCCGCAGCAGAGACCAAGCCACTGA 691
QY 877 GGAACCTGCGGAGCTGGAGAGAGAGAGCGAGCCCGCTCGGTTCAGAGGTGACAGCG 936
DB 692 GGAACCTGCGGAGCTGGAGAGAGAGAGCGAGCCCGCTCGGTTCAGAGGTGACAGCG 751
QY 937 CTGGAAGAGTCTGGCTGCTGCAACATTTGAGTATGAGAACTCGAGAGAGAGAAACA 996
DB 752 CTGGAAGAGTCTGGCTGCTGCAACATTTGAGTATGAGAACTCGAGAGAGAGAAACA 811
QY 997 CTACTTATGATTCGCGCGGATGAAAACTACAGTCCCAATTTGGGCTGCTCCCTTC 1056
DB 812 CTACTTATGATTCGCGCGGATGAAAACTACAGTCCCAATTTGGGCTGCTCCCTTC 871
QY 1057 CGTGGCCCAAGAGTCAAGACTGCCCCCGGACCTCCCAATTCCTGATTCGCTCCCA 1116
DB 872 CGTGGCCCAAGAGTCAAGACTGCCCCCGGACCTCCCAATTCCTGATTCGCTCCCA 931
QY 1117 TGAACCCGAAGCCATTCACATCCCAACGAAAGCCCAAGGCTGAGCCGGCTCTT 1176
DB 932 TGAACCCGAAGCCATTCACATCCCAACGAAAGCCCAAGGCTGAGCCGGCTCTT 991
QY 1177 CCACTTCTTGAACACCCCAATGCGAAGGTCTCAGAGCTCCAGCAACGGCTTCGGGGAC 1236
DB 992 CCACTTCTTGAACACCCCAATGCGAAGGTCTCAGAGCTCCAGCAACGGCTTCGGGGAC 1051

QY 1237 CCAAGACGGGAGAGACATTTGTAGAGTCCCCCAAGGCCGAGGAGAGTGTGGTGT 1296
DB 1052 CCAAGACGGGAGAGACATTTGTAGAGTCCCCCAAGGCCGAGGAGAGTGTGGTGT 1111
QY 1297 GGGCCAGTGTGCGCAGAGGGGCAAGAAACAAGCCCTCTTGGAGACCGGCATTCCTGCGGT 1356
DB 1112 GGGCCAGTGTGCGCAGAGGGGCAAGAAACAAGCCCTCTTGGAGACCGGCATTCCTGCGGT 1171
QY 1357 GTCCCTCTCCGCCCACTTGGCTGCGAGCCCGGCTCTCTCCCTCCCTAGCCCTCCG 1416
DB 1172 GTCCCTCTCCGCCCACTTGGCTGCGAGCCCGGCTCTCTCCCTCCCTAGCCCTCCG 1231
QY 1417 GCACAAAGACCAAGACACGAGCCCAAGAGAGACAGAGGCTGCGGAGCTGACAGGC 1476
DB 1232 GCACAAAGACCAAGACACGAGCCCAAGAGAGACAGAGGCTGCGGAGCTGACAGGC 1291
QY 1477 ACCACTGGCTCAGTGTGCTGTCTGTGGGGCGAGACCTTGGGAGCTGCCGCTT 1536
DB 1292 ACCACTGGCTCAGTGTGCTGTCTGTGGGGCGAGACCTTGGGAGCTGCCGCTT 1351
QY 1537 GGTGTGTATGAGAGCCAGGCGGGGACCGGTCCAGAGACATGAGACCAACACCA 1596
DB 1352 GGTGTGTATGAGAGCCAGGCGGGGACCGGTCCAGAGACATGAGACCAACACCA 1411
QY 1597 TGAACATCACCACTTACCACTTCTACAGACA 1633
DB 1412 TGAACATCACCACTTACCACTTCTACAGACA 1448

RESULT 4
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Query Match 42.1%; Score 783; DB 10; Length 2379;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 GGCTCCCTTCGCTGCGCAAGTGAAGCTGCCCCCGCAGCTTCAATCCACTCGA 1105
DB 840 GGCTCCCTTCGCTGCGCAAGTGAAGCTGCCCCCGCAGCTTCAATCCACTCGA 899
QY 1106 TCTGCTCCATGAGCCGGAAGCATTCACATCCCAACGAAAGCCCAAGGGGTGAG 1165
DB 900 TCTGCTCCATGAGCCGGAAGCATTCACATCCCAACGAAAGCCCAAGGGGTGAG 959
QY 1166 CCGGCTCTTTCACATTCCTTGAACCCCAATGCGCAAGGTCTCAGAGCTCCAGCAACGG 1225
DB 960 CCGGCTCTTTCACATTCCTTGAACCCCAATGCGCAAGGTCTCAGAGCTCCAGCAACGG 1019
QY 1226 CTCGGGGGACCAAGAGCGGAGCAAGCACTTTGTGAGTCCCCCAAGGCCAGGCGAAG 1285
DB 1020 CTCGGGGGACCAAGAGCGGAGCAAGCACTTTGTGAGTCCCCCAAGGCCAGGCGAAG 1079
QY 1286 AGTGTGTGTGGGCGCAGTGTGCGCAAGAGGGGCAAGAAACAAGCCCTCTGGGACCGGCC 1345
DB 1080 AGTGTGTGTGGGCGCAGTGTGCGCAAGAGGGGCAAGAAACAAGCCCTCTGGGACCGGCC 1139
QY 1346 ATCCCTCGGATGTCCCTCCGCCCACTGCTGCGAGCCCGGCTCTTCCCTCTA 1405

Db	1140	ATCCCTGGGGTGTCCCTCCGCCACCTGGGTGCAACCCGGCCCTCTCCCTCCCTA	1199
QY	1406	GCCCCCTCGGGCACAGAGCACAAGCAGAGCCAGAGAGAGCCAGCAGGGCTGCCGG	1465
Db	1300	GCCCCCTCGGGCACAGAGCACAAGCAGAGCCAGAGCCAGAGAGCCAGGGCTGCCGG	1259
QY	1466	GGCCGTGAGGACCACTGGGCTCAGGTGGCCCTGTCTCGGGGGGGGACCACTGGGGAG	1525
Db	1260	GGCTTGACGGACCACTGGCTCAGGTGGCCCTGTCTCGGGGGGGGACCACTGGGGAG	1319
QY	1526	CTGCCCGCTTGTGTGTATGAGAGCCAGGCGGGCAGCCGGTCAAGACATGAGCAC	1588
Db	1320	CTGCCCGCTTGTGTGTATGAGAGCCAGGCGGGCAGCCGGTCAAGACATGAGCAC	1379
QY	1386	CACCAACCAATGAACATCACCACTTAACAACCTTCTACAGACATAGAGCCCTCC	1645
Db	1380	CACCAACCAATGAACATCACCACTTAACAACCTTCTACAGACATAGAGCCCTCC	1439
QY	1646	CCAGGGCCCCACCCGCGCATATGAAGAGCCCAACCCCGGACACCAAGGATTAATTT	1705
Db	1440	CCAGGGCCCCACCCGCGCATATGAAGAGCCCAACCCCGGACACCAAGGATTAATTT	1499
QY	1706	CTATTAAATTAATGTTATATGATGATTAATTTGTTAATTAATTAATTAATTAATTAAT	1765
Db	1500	CTATTAAATTAATGTTATATGATGATTAATTTGTTAATTAATTAATTAATTAATTAAT	1559
QY	1766	ATTTAAGCTACCTACATGTAGAGATCTATGGAAAACAGAACTTAACCTTTATTTATAT	1825
Db	1560	ATTTAAGCTACCTACATGTAGAGATCTATGGAAAACAGAACTTAACCTTTATTTATAT	1619
QY	1826	GTT 1828	
Db	1620	GTT 1622	

```

RESULT 5
US-09-764-891-5904
/ Sequence 5904, Application US/09764891
/ Publication No. US2003007808A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC006
/ CURRENT APPLICATION NUMBER: US/09/764, 891
/ CURRENT FILING DATE: 2001-01-17
/ Prior application data removed - consult PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 10231
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5904
/ LENGTH: 2379
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-764-891-5904

```

Query Match	42.1%	Score 783	DB 10	Length 2379
Best Local Similarity	100.0%	Pred. NC	0	
Matches 783	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

[illegible]

Db	1020	CTCCGGGGCAACCGAGACGGGAGCAACAATTGTGAGGTCCCCCAAGGCGCGCAAG	1079
OY	1286	AGTGTGGTGTGGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGCC	1345
Db	1080	AGTGTGGTGTGGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGCC	1139
OY	1346	ATCCCTGGGGTGTCCCTCCGCCCACTGGCTGCAGCCCGGCGCTCTCCCTCCCTTA	1405
Db	1140	ATCCCTGGGGTGTCCCTCCGCCCACTGGCTGCAGCCCGGCGCTCTCCCTCCCTTA	1139
OY	1406	GCCCCCTCGGGCACAAGAGCACAGCACTGAGCCCAAGAGAGCCAGCAGGCTGGCCGG	1465
Db	1200	GCCCCCTCGGGCACAAGAGCACAGCACTGAGCCCAAGAGAGCCAGCAGGCTGGCCGG	1259
OY	1466	GGCCCTGACGACACACATGGCCCTCAGGGTGGCCCTGTCTGGGGGGGGGACCTGGGGGAG	1525
Db	1260	GGCCCTGACGACACACATGGCCCTCAGGGTGGCCCTGTCTGGGGGGGGGACCTGGGGGAG	1319
OY	1526	CTGCCCGGCTGTGGTGTATGAGAGCGAGGCCGGGAGCGGTCCAGAGACATGAGCAC	1585
Db	1320	CTGCCCGGCTGTGGTGTATGAGAGCGAGGCCGGGAGCGGTCCAGAGACATGAGCAC	1379
OY	1586	CACCAACAACATGAACATCACCACTTACCACTTCTTACAGACATGAGGCCCTCC	1645
Db	1380	CACCAACAACATGAACATCACCACTTACCACTTCTTACAGACATGAGGCCCTCC	1439
OY	1646	CCAGGGGCCCCCGCTGCATATGAAAGACCCCAACCCCGGACACCAACAAGCATTTATT	1705
Db	1440	CCAGGGGCCCCCGCTGCATATGAAAGACCCCAACCCCGGACACCAAGGCATTTATT	1499
OY	1706	CTATTAAATTATTTATTAATGATGATTTATTTATTAATTAATTTATTTATCTCCATAT	1765
Db	1500	CTATTAAATTATTTATTAATGATGATTTATTTATTAATTAATTTATTTATCTCCATAT	1559
OY	1766	ATTTAAGCTAGCCTACATGTAGAGATCTATGAAACAAGAACTTAACTTTATTTAT	1825
Db	1560	ATTTAAGCTAGCCTACATGTAGAGATCTATGAAACAAGAACTTAACTTTATTTAT	1619
OY	1826	GTT 1828	
Db	1620	GTT 1622	

```

RESULT 6
US-10-087-192-178
/ Sequence 178, Application US/10087192
/ Publication No. US20020182586A1
/ GENERAL INFORMATION:
/ APPLICANT: Morris, David W.
/ APPLICANT: Engelhard, Eric K.
/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
/ TITLE OF INVENTION: CANCER
/ FILE REFERENCE: 529452000122
/ CURRENT APPLICATION NUMBER: US/10/087,192
/ CURRENT FILING DATE: 2002-03-01
/ PRIOR APPLICATION NUMBER: US 09/747,377
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 09/798,586
/ PRIOR FILING DATE: 2001-03-02
/ NUMBER OF SEQ ID NOS: 2059
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 178
/ LENGTH: 106315
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)-(106315)
/ OTHER INFORMATION: n = A,T,C or G
US-10-087-192-178

```

Query Match	42.1%;	Score 783;	DB 13;	Length 106315;
Best Local Similarity	100.0%;	Pred. No. 0;		

	Matches	783;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1046	GGCTCCCTTCCTCGTGC	CAAGAGTGAACTGCCCCCCC	CGACCTC	CATCCACTCGA	1105				
Db	95728	GGCTCCCTTCCTCGTGC	CAAGAGTGAACTGCCCCCCC	CGACCTC	CATCCACTCGA	95787				
Qy	1106	TTCGTGCCATGAGCGGGA	GCATTCACATCCCAACCGAAAGCCCAAGCGGTGAC	1165						
Db	95788	TTCGTGCCATGAGCGGGA	GCATTCACATCCCAACCGAAAGCCCAAGCGGTGAC	95847						
Qy	1166	CCGGCTCTCTTCCACTTT	GACACCACCCAAAGTCTCAGA	GTCCAGCAACGG	1225					
Db	95848	CCGGCTCTCTTCCACTTT	GACACCACCCAAAGTCTCAGA	GTCCAGCAACGG	95907					
Qy	1226	CTCCGGGGCACCCAGAG	CGGAGCAACTTTGTAGTCCCC	CAAAGCCCAAGGCAAG	1285					
Db	95908	CTCCGGGGCACCCAGAG	CGGAGCAACTTTGTAGTCCCC	CAAAGCCCAAGGCAAG	95967					
Qy	1286	AGTGTGGGTGTGGGCCA	CGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGGC	1345						
Db	95968	AGTGTGGGTGTGGGCCA	CGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGGC	96027						
Qy	1346	ATCCCTCGGTGTGCCCT	CCGCGCACCTTGCTGACAGCCCGGCTCTCCCTCCCTTA	1405						
Db	96028	ATCCCTCGGTGTGCCCT	CCGCGCACCTTGCTGACAGCCCGGCTCTCCCTCCCTTA	96087						
Qy	1406	GCCCCCTCTCGGGCA	CAAGAACCAAGACCCAGCCAAGAGAGCCAGACGGCTGCGG	1465						
Db	96088	GCCCCCTCTCGGGCA	CAAGAACCAAGACCCAGCCAAGAGAGCCAGACGGCTGCGG	96147						
Qy	1466	GGCTTGAGGACCACTG	CGCTCAGGTGGCCCTGTCTGGGGGGGAGACCTGCGGGAG	1525						
Db	96148	GGCTTGAGGACCACTG	CGCTCAGGTGGCCCTGTCTGGGGGGGAGACCTGCGGGAG	96207						
Qy	1526	CTGCCCGCTTGTGGTGT	GTATGAGAGCCAGGCGGGCAGCGGTCTCAGAGACATGAGCAC	1585						
Db	96208	CTGCCCGCTTGTGGTGT	GTATGAGAGCCAGGCGGGCAGCGGTCTCAGAGACATGAGCAC	96267						
Qy	1586	CACCACCAACCATGAA	CAACACCACTTACACACCTTCTACGACATAGAGCCCTCC	1645						
Db	96268	CACCACCAACCATGAA	CAACACCACTTACACACCTTCTACGACATAGAGCCCTCC	96327						
Qy	1646	CCAGGGCCCCACCTG	CGCATATGAAAGACCCCGACACCAAGGCATTATTAT	1705						
Db	96328	CCAGGGCCCCACCTG	CGCATATGAAAGACCCCGACACCAAGGCATTATTAT	96387						
Qy	1706	CTATTAAATTATGTTAT	TATGATGATTAATTATTAATAATTAATTGTTACTCACTAAT	1765						
Db	96388	CTATTAAATTATGTTAT	TATGATGATTAATTATTAATAATTAATTGTTACTCACTAAT	96447						
Qy	1766	ATTTAGCTAGCTACAT	GTAGAGATCTATGAAAACACAGAACTAAACTTTTATTATAT	1825						
Db	96448	ATTTAGCTAGCTACAT	GTAGAGATCTATGAAAACACAGAACTAAACTTTTATTATAT	96507						
Qy	1826	GTTT	1828							
Db	96508	GTTT	1828							
RESULT 7										
US-09-993-966-21										
; Sequence 21, Application US/09993966										
; Publication No. US20030186232A1										
; GENERAL INFORMATION:										
APPLICANT: ROHAN, MICHAEL										
TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,										
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF										
FILE REFERENCE: 014024/028073										
CURRENT APPLICATION NUMBER: US/09/993, 966										
CURRENT FILING DATE: 2001-11-27										
PRIOR APPLICATION NUMBER: 60/252, 884										
PRIOR FILING DATE: 2000-11-27										
PRIOR APPLICATION NUMBER: 60/291, 109										

```

; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-21

Query Match      31.7%; Score 590; DB 10; Length 590;
Best Local Similarity 100.0%; Pred. No. 1,9e-273;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1047  GCTCCCTTCCTCGTGGCCCAAGAGTCAAGACTGCCCCCGCACTCCCAATCCACTCGAT 1106
DB      1    GCTCCCTTCCTCGTGGCCCAAGAGTCAAGACTGCCCCCGCACTCCCAATCCACTCGAT 60

QY      1107  CTGCGCTCCCAATGAGCCGGAGAGCCATTCACATCCCAACCCGAAGCCCAAGGGGTGAGC 1166
DB      61    CTGCGCTCCCAATGAGCCGGAGAGCCATTCACATCCCAACCCGAAGCCCAAGGGGTGAGC 120

QY      1167  CGGCGCTCCTTCACATTGCTTGAACACCCCAATTCGCAAGTCTCAGAGCTCCAGCAAGC 1226
DB      121  CGGCGCTCCTTCACATTGCTTGAACACCCCAATTCGCAAGTCTCAGAGCTCCAGCAAGC 180

QY      1227  TCCGGGGCACCAGAGACGGAGCAAGACATTGTGAGGTGCCCAAGGCCCAAGGGCAAGA 1286
DB      181  TCCGGGGCACCAGAGACGGAGCAAGACATTGTGAGGTGCCCAAGGCCCAAGGGCAAGA 240

QY      1287  GTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAGCCCCCTGAGGACCCGCA 1346
DB      241  GTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAGCCCCCTGAGGACCCGCA 300

QY      1347  TCCCTGCGGAGTATCCCCCTCGCGCCCAAGCTGAGCTGCGACCGCGGCTCTCCCTCCCTAG 1406
DB      301  TCCCTGCGGAGTATCCCCCTCGCGCCCAAGCTGAGCTGCGACCGCGGCTCTCCCTCCCTAG 360

QY      1407  CCCCCCTCGGCAACAAGAGCAAGACCCGAGCAAGAGAGGCAAGGAGCTGCGGG 1466
DB      361  CCCCCCTCGGCAACAAGAGCAAGACCCGAGCAAGAGAGGCAAGGAGGCTGCGGG 420

QY      1467  GCTGTGAGGCAACCACTGGCTCTAGGTGGCTGTCTCTGGGGCGGGAGCACTTGGGGAGC 1526
DB      421  GCTGTGAGGCAACCACTGGCTCTAGGTGGCTGTCTCTGGGGCGGGAGCACTTGGGGAGC 480

QY      1527  TGCCCGGCTTGTGTGTATAGAGCCAGGCGGGGACCGGCTCCAGAGCATGAGCACCC 1586
DB      481  TGCCCGGCTTGTGTGTATAGAGCCAGGCGGGGACCGGCTCCAGAGCATATAGCACCC 540

QY      1587  ACCACCAACATGAAATCAACCAACCATTAACAACAATTCTAACAGACATAG 1636
DB      541  ACCACCAACATGAAATCAACCAACCATTAACAACAATTCTAACAGACATAG 590

RESULT 8
US-09-918-995-8686
; Sequence 8686, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8686
; LENGTH: 427

```



```
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO CHR16.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
;; OTHER INFORMATION: SWISSPROT HIT: Q23977, EVALUE 7.70e-02
;; OTHER INFORMATION: NT HIT: g114916432, EVALUE 9.00e-71
;; OTHER INFORMATION: EST_HUMAN HIT: BF920913.1, EVALUE 3.00e-68
US-10-029-386-7990
```

Query Match 7.4%; Score 137; DB 15; Length 598;

Best Local Similarity 100.0%; Pred. No. 2.2e-55;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 280 AGGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGG 339
DB 228 AGGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGCGATCGAGAGTGATCGG 287
QY 340 GAGACAGCGCTGCGCCGGGGCGGTGTCTGGGAGCCCGACAGCTGGGTTGGCGGGCAACAT 399
DB 288 GAGACAGCGCTGCGCCGGGGCGGTGTCTGGGAGCCCGACAGCTGGGTTGGCGGGCAACAT 347
QY 400 AGGCCGGAAGCACCCCGG 416
DB 348 AGGCCGGAAGCACCCCGG 364
```

RESULT 12
US-10-016-157A-126

; Sequence 126, Application US/10016157A

; Publication No. US20020192220A1

; GENERAL INFORMATION:

; APPLICANT: Sun, Yongming

; APPLICANT: Recipon, Hervé

; APPLICANT: Ghosh, Malayika

; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro

; FILE REFERENCE: DEX-0253

; CURRENT APPLICATION NUMBER: US/10/016,157A

; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,717

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 250

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 126

; LENGTH: 1950

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-016-157A-126

Query Match 7.4%; Score 137; DB 13; Length 1950;

Best Local Similarity 100.0%; Pred. No. 1.9e-55;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 281 GGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGGG 340
DB 838 GGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGGG 897
QY 341 AGACAGCGCTGCGCCGGGGCGGTGTCTCGGAGCCCGACAGCTGCGGTTGGCGGGCAACATA 400
DB 898 AGACAGCGCTGCGCCGGGGCGGTGTCTCGGAGCCCGACAGCTGCGGTTGGCGGGCAACATA 957
QY 401 GCGCGAAGCACCCCGGGA 417
DB 958 GCGCGAAGCACCCCGGGA 974
```

US-10-029-386-21690

Query Match	7.2%	Score 133;	DB 15;	Length 133;
Best Local Similarity	100.0%	Pred. No. 2.2e-53;		
Matches 133; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 283 TGACAGCTTCGCGCGTAGCGCTGCCTCGGAGGGGCATTCAGAGAGTGCATCGGGAG 342

07 ACACGCGTGCCTGGGGGGTGTCTGGGACCCCGACAGCTGCGGTGGCGGGGACCAATAGG 402
343 ACACGCGTGCCTGGGGGGTGTCTGGGACCCCGACAGCTGCGGTGGCGGGGACCAATAGG 402
Db 61 ACACGCGTGCCTGGGGGGTGTCTGGGACCCCGACAGCTGCGGTGGCGGGGACCAATAGG 120

Qy	403	CCGAAGCACC	CGG	415
Db	121	CCGAAGCACC	CGG	133

RESULT 15
ME-09-093

US-09-993-966-20

Sequence 20, Application US/09993966

Publication No. US20030186232A1

ADDITIONAL INFORMATION: ; GENERAL INFORMATION:

APPLICANT: ROHAN, MICHAEL

1. TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,

1. TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF

FILE REFERENCE: 014024/0280733

CURRENT APPLICATION NUMBER: US/09/993,966

CURRENT FILING DATE: 2001-11-27

PRIOR APPLICATION NUMBER: 60/252,884

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/291,109
PRIORITY DATE: 2001-07-15

PRIOR FILING DATE: 2001-05-16

PRIOR APPLICATION NUMBER: 60/325,571

PRIOR FILING DATE: 2001-10-
NUMBER OF SEQ. TO NO. 36

NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30; SEQ ID NO 20
LENGTH: 139; LENGTH: 1
TYPE: DNA

TYPE: DNA

ORGANISM: HO

Query Match	6.9%	Score 128	DB 10	Length 128
-------------	------	-----------	-------	------------

```

Matches      128; Conservative    0; Mismatches    0; Indels      0; Gaps        0;

```

QY 919 GTTCAGGGGACAGACCGCCCTGGAGAGCATCTGGCTGCTACCAACCATTTGGCTGATGATGAA 978

Db 1 GTTCAGGGGACAGACCGCCCTGGAGAGCATCTGGCTGCTACCAACCATTTGGCTGATGATGAA 60

Qy	979	1038
CATGAGAGGAGAAACCACTACTTATGATCTCGCCGGATAGAAAACTACAGTCCCAATT		
Db	61	120
CATGAGAGGAGAAACCACTACTTATGATCTCGCCGGATAGAAAACTACAGTCCCAATT		

Qy	1039	TGGGCTG	1046
Db	121	TGGGCTG	128

Search completed: December 30, 2004, 03:48:42
Job time : 1491 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 18:07:39 ; Search time 12706 Seconds
(Without alignments)
6918.901 Million cell updates/sec

Title: US-09-993-966-5

Sequence: 1859
1 gaattcgccttcataacg.....aaaaaaaaaagcgcgccgc 1859

Scoring table: OLIGO₂NUC₂
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmb1.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sgs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1786	96.1	1786	AY061883	AY061883 Homo sapi
2	1763	94.8	2604	BC051288	BC051288 Homo sapi
3	1610	86.6	1788	AB062886	AB062886 Homo sapi
4	1438	77.4	1438	AF358135	AF358135 Homo sapi
5	1336	71.9	1438	AR405917	AR405917 Sequence
6	812	43.7	1053	CQ721681	CQ721681 Sequence
7	783	42.1	163319	HSAS03140	HSAS03140 Homo sapi
8	783	42.1	169773	AC007608	AC007608 Homo sapi
9	783	42.1	170791	AC007334	AC007334 Homo sapi
10	528	28.4	178790	AC145238	AC145238 Pan trogl
11	528	28.4	207486	AC145175	AC145175 Pan trogl
12	137	7.4	1950	AX557538	AX557538 Sequence
13	136	7.3	2340	CQ741468	CQ741468 Sequence
14	124	6.7	194869	AC145003	AC145003 Papio anu
15	124	6.7	198189	AC145043	AC145043 Papio anu
16	70	3.8	181401	AC146663	AC146663 Sus scrofa
17	68	3.7	2469	CO843478	CO843478 Sequence
18	68	3.7	2469	AK124375	AK124375 Homo sapi
19	65	3.5	199277	AC145445	AC145445 Canis fam

20	58	3.1	670	6	AX530283	AX530283 Sequence
21	57	3.1	417	6	AX530282	AX530282 Sequence
22	57	3.1	510	6	AX530281	AX530281 Sequence
23	56	3.0	636	6	AX530289	AX530289 Sequence
24	56	3.0	651	6	AX530272	AX530272 Sequence
25	56	3.0	686	6	AX530282	AX530282 Sequence
26	56	3.0	2048	4	AJ715983	AJ715983 Equus cab
27	56	3.0	3951	10	AF304853	AF304853 Rattus no
28	56	3.0	208061	2	AC145254	AC145254 Bos tauru
29	55	3.0	547	6	AX530274	AX530274 Sequence
30	55	3.0	1245	6	AX530269	AX530269 Sequence
31	54	2.9	1874	6	AX528996	AX528996 Sequence
32	54	2.9	1874	6	AX529000	AX529000 Sequence
33	54	2.9	2437	3	AJ620505	AJ620505 Lacanobia
34	54	2.9	2461	8	MSY428052	AJ428052 Malassezi
35	54	2.9	2605	6	AX529022	AX529022 Sequence
36	53	2.9	699	3	SFR488181	AJ488181 Spodopter
37	53	2.9	711	8	VAL457214	AJ457214 Viacum al
38	53	2.9	1434	6	CQ792377	CQ792377 Sequence
39	53	2.9	1434	6	AX456458	AX456458 Sequence
40	53	2.9	3003	10	AB090950	AB090950 Cavia por
41	53	2.9	3281	6	AX708831	AX708831 Sequence
42	52	2.8	577	6	AX530278	AX530278 Sequence
43	51	2.7	2241	5	SSA56995	AJ56995 Salmo sal
44	50	2.7	226	6	AX530284	AX530284 Sequence
45	50	2.7	1401	6	AX172351	AX172351 Sequence

ALIGNMENTS

RESULT 1
LOCUS AY061883
DEFINITION Homo sapiens naked protein (NKD)
ACCESSION AY061883
VERSION AY061883.1
KEYWORDS GI:17978536
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1786)
AUTHORS Yan, D., Wiesmann, M., Rohan, M., Chan, V., Jefferson, A. B., Guo, L., Sakamoto, D., Caothien, R. H., Fuller, J. H., Reinhard, C., Garcia, P. D., Randazzo, F. M., Escobedo, J., Fanli, W. J., and Williams, L. T.
Elevated expression of axin2 and nkhd mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14873-14978 (2001)

JOURNAL PUBMED 11752446
REFERENCE 2 (bases 1 to 1786)
AUTHORS Rohan, M., Chan, V., Yan, D., Fanli, W. J., and Williams, L. T.
Direct Submission
Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA
JOURNAL TITLE
FEATURES
source
1. 1786
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. 1786
/gene="NKD"
159. 1571
/gene="NKD"
/note="HNKD"
/codon_start=1
/product="naked protein"
/protein_id="AAU32374.1"
/db_xref="GI:17978537"
/translation="MGKLSKPAACKRRSPGDSFAVSAAMARKGIEWIGRQCP
GGVSGPRLACTIGTSTRELVDVLRDTLSEEDDFLEVALPEKTDGSGDGE
KMERVSEPCGSKOLKPEELCDVMSDEDSOEWFETLYDDNNKGVREDTLSL
HTIYEVDSSVNSPSSKMLRVKLTVAAPDGSGSKSVLVNQADLQSAARAEKPKPE

DIRSEKKORAPLRFOGDSRLRLEOSGCVHHCVNDENIERRNHYLDLAGIENYISORFPGS
PSYAKSELPRPTNSRSHBEPAIHIHPRKPOGDPASPHLDPIAKVSELOR
LRYODDSKHVRSPKAGKSVGVGHARGNRPNRPIDRAIPANPSPHILASALLP
SLAPLHKHKHRAKESQOGCKRGLADPLASGSPVLRHLELPAIVTESQAGQPVQ
RHEHHHHHHHHHFFYT"

ORIGIN

Query Match 96.1%; Score 1786; DB 9; Length 1786;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 66 AGTCGGGCGCGGCGAGCGGCGAGAGCGGCTCCGCGCGCGCTCGGCTCGGCTCGG 125
Db 1 AGTCGGGCGCGGCGAGCGGCGAGAGCGGCTCCGCGCGCGCTCGGCTCGGCTCGG 60
QY 126 CTGGGGGGCTCTCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
Db 61 CTGGGGGGCTCTCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 186 CATGGCTTAGGGAGGCTCCGGCGCGCGAGCGGCGAGAGAGAGAGAGAGAGAGAGAG 245
Db 121 CATGGCTTAGGGAGGCTCCGGCGCGCGAGCGGCGAGAGAGAGAGAGAGAGAGAGAG 180
QY 246 CGGCGCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
Db 181 CGGCGCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 306 CTTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
Db 241 CTTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 366 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
Db 301 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 426 GCGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
Db 361 GCGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 486 CCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
Db 421 CCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 546 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
Db 481 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 606 TGTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
Db 541 TGTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 666 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
Db 601 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 726 ACTCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
Db 661 ACTCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 786 CCCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 845
Db 721 CCCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 846 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905
Db 781 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 906 GAGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
Db 841 GAGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 966 GCGTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
Db 901 GCGTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 1026 ACAGGTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAG 1085
Db 961 ACAGGTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAG 1020
QY 1086 GCACCTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAGAG 1145
Db 1021 GCACCTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAGAG 1080
QY 1146 GAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1205
Db 1081 GAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1206 TCTTAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265
Db 1141 TCTTAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1266 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
Db 1201 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1326 AGCCCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385
Db 1261 AGCCCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1386 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
Db 1321 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
QY 1446 AGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
Db 1381 AGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1506 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565
Db 1441 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1566 CGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
Db 1501 CGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1626 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685
Db 1561 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1686 CACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1745
Db 1621 CACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1746 ATTATTGTACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Db 1681 ATTATTGTACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1806 AACTTAACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1851
Db 1741 AACTTAACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1786
```

```
Db 901 GCGTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 1026 ACAGGTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAG 1085
Db 961 ACAGGTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAG 1020
QY 1086 GCACCTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAGAG 1145
Db 1021 GCACCTCCCAATTTGGGCTGGCTCCCTTCCGCGGCGAGAGAGAGAGAGAGAGAGAG 1080
QY 1146 GAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1205
Db 1081 GAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1206 TCTTAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265
Db 1141 TCTTAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1266 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
Db 1201 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1326 AGCCCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385
Db 1261 AGCCCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1386 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
Db 1321 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
QY 1446 AGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
Db 1381 AGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1506 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565
Db 1441 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1566 CGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
Db 1501 CGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1626 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685
Db 1561 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1686 CACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1745
Db 1621 CACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1746 ATTATTGTACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Db 1681 ATTATTGTACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1806 AACTTAACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1851
Db 1741 AACTTAACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1786
```

RESULT 2
BC051288 2604 bp mRNA linear PRI 30-JUN-2004
LOCUS BC051288
DEFINITION Homo sapiens naked cuticle homolog 1 (Drosophila), mRNA (cDNA clone
WGC:59666 IMAGE:6646089), complete cds.
ACCESSION BC051288
VERSION BC051288.1 GI:30410965
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2604)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

TITLE	JOURNAL PUBMED REFERENCE AUTHORS TITLE JOURNAL	REMARK COMMENT
Klausner R.D., Collins F.S., Wagner L., Shennem C.M., Schuler G.D. Altschul S.F., Zeeberg B., Butlov K.H., Schaefer C.F., Bat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hatch F., Datchenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueding T.B., Toshitsuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H., Richards S.W., Moyley K.C., Hale S., Garcia A.M., Gay L.J., Hilly K.S., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J.J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Boffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butcherfield Y.S., Krzywinski M.I., Skalka U., Smalins D.E., Schnerch A., Schein J.R., Jones S.J. and Marra M.A.	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences <i>Proc. Natl. Acad. Sci. U.S.A.</i> 99 (26), 16899-16903 (2002)	
2 (bases 1 to 2604) 12477932	Strausberg R. Direct Submission Submitted (23-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.linl.gov>
Series: IRAL Plate: 47 Row: 0 Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20127660

FEATURES	source	Location/Qualifiers
gene		1..2604
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="MGC:59666 IMAGE:664608"
		/tissue_type="Ovary, teratocarcinoma"
		/clone_id="NH MGC_109"
		/lab_host="DH10B-R"
		/note="Vector: pOTB7"
		1..2604
		/gene="NKD1"
		/db_xref="locusID:85407"
		/db_xref="MIM:607851"
		225..1637
		/gene="NKD1"
		/codon_start=1
		/product="naked cuticle homolog 1"
		/protein_id="AAH51288.1"
		/db_xref="GI:30410966"
CDS		

```

/ab xref="LocusID:85407"
/db xref="MIM:607851"
/translation="MCKLHKKPAAVCCRRSPSGDSSPAVSAAMARKIEEMIGRCRCP
GVSGSPRQRLACTIGTSELRELVDGVRDLTISEEDDFLEVLPEKTDGLDSSGGE
KNHREVSPEPCGSRKKQLKPELQCDVSMEDSSQEWFTLYEPFNMKVREDDTSL
HTLYEVVDSVSNHPTSSKMLRYLAIVAPGSGSKSLVNOQDLGARPAETKPRE
DLSEWKKORAPLRFGQSDRLTSGCGCHVDNIBERNRYLDLAGIENTSOGCPDS
PSVAQGSSELPPRTSNPTFRKSHREPAIHITHRKPGQYDPAFSFHTDPIKAVSELQOR
LRTOGSGKRIFVRSPPKAQGSVSGVGHANARANKPPIGAIIPAVSPAHILASAPLL
SLAPLGHKKHKKHAKESQCGCLOAPRGVGLGRENHRELPALVAVYESQAQCPQ
RHHHHHHHHHHHHFYQ"

```

[illegible]

```
QY 906 GAGCCCCCTCAGGTTCCAGGTTGACAGCCGCTGGAGCAGTCTGCTGCTACCAACATT 965
DB 907 GAGCCCCCTCAGGTTCCAGGTTGACAGCCGCTGGAGCAGTCTGCTGCTACCAACATT 966
QY 966 GCGTAGATGAAACATCGAGAGGAGAAACCACTATTAGATCTCGCGGAGTAGAAAAC 1025
DB 967 GCGTAGATGAAACATCGAGAGGAGAAACCACTATTAGATCTCGCGGAGTAGAAAAC 1026
QY 1026 ACAAGTCCCAATTTGGGCTGGCTCCCTTCCGTCGGGCGCCAGAGTCAAGACTGCCCC 1085
DB 1027 ACAAGTCCCAATTTGGGCTGGCTCCCTTCCGTCGGGCGCCAGAGTCAAGACTGCC 1086
QY 1086 GCACCTCCATCCCATCTGATCTGCTGCCATGAGCGGAGACCATCCATCCACACC 1145
DB 1087 GCACCTCCATCCCATCTGATCTGCTGCCATGAGCGGAGACCATCCATCCACACC 1146
QY 1146 GAAAGCCCCAAGGCGTGGAGCCCGGCTCTTCCACTTCTTGAACCCCAATCGCCAA 1205
DB 1147 GAAAGCCCCAAGGCGTGGAGCCCGGCTCTTCCACTTCTTGAACCCCAATCGCCAA 1206
QY 1206 TCTCAGAGCTTCAGCAACGGCTCTCGGGGACCCAGAGCGGAGACGACACTTTGTAG 1265
DB 1207 TCTCAGAGCTTCAGCAACGGCTCTCGGGGACCCAGAGCGGAGACGACACTTTGTAG 1266
QY 1266 CCCCCAAGGCGCGAGGAGAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAA 1325
DB 1267 CCCCCAAGGCGCGAGGAGAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAA 1326
QY 1326 AGCCCCCTCTGGGACCCGCACTCCCTGCGTGTCTCCCTCCGCCCACTGGCTGCAAG 1385
DB 1327 AGCCCCCTCTGGGACCCGCACTCCCTGCGTGTCTCCCTCCGCCCACTGGCTGCAAG 1386
QY 1386 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
DB 1387 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1446
QY 1446 AGAGCCAGCAGGAGGCTGCGGGGAGCTGAGAGCACCAGTGGCTGAGGCGCTGCTG 1505
DB 1447 AGAGCCAGCAGGAGGCTGCGGGGAGCTGAGAGCACCAGTGGCTGAGGCGCTGCTG 1506
QY 1506 GCGGGAGAGCCTGCGGGAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1565
DB 1507 GCGGGAGAGCCTGCGGGAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1566
QY 1566 CCGTCCAGAGACATGAGACACACACACACATGAACATCAACCATTAACCACTTTCT 1625
DB 1567 CCGTCCAGAGACATGAGACACACACACACATGAACATCAACCATTAACCACTTTCT 1626
QY 1626 ACAGACATAGAGCCCTCCCAAGGGCCCACTGTCATATGAAAGACCCCAACCCG 1685
DB 1627 ACAGACATAGAGCCCTCCCAAGGGCCCACTGTCATATGAAAGACCCCAACCCG 1686
QY 1686 CACCAAGGCAATTTATTTATTAATTAATTTATTTATTTATTTATTTATTTATA 1745
DB 1687 CACCAAGGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1746
QY 1746 ATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1805
DB 1747 ATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1806
QY 1806 AACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1828
DB 1807 AACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1829
```

```
RESULT 3
LOCUS AB062886 1788 bp mRNA linear PRI 20-OCT-2001
DEFINITION Homo sapiens mRNA for Dvl-binding protein NKD1, complete cds.
ACCESSION AB062886
VERSION AB062886.1 GI:16303259
KEYWORDS
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Butheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Molecular cloning, gene structure, and expression analyses of NKD1
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE Int. J. Oncol. 19 (5), 963-969 (2001)
PUBMED 21490203
REFERENCE 11604995
AUTHORS 2 (bases 1 to 1788)
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center
Research Institute, Genetics and Cell Biology Section, Tsukiji
5-chome, Chuo-ku, Tokyo 104-0045, Japan (E-mail:mkatoh@ncc.go.jp,
Tel:81-3-3542-2511)
FEATURES
source
1..1788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="human fetal tissues"
1..1788
/gene="NKD1"
5..1418
/gene="NKD1"
/codon_start=1
/product="Dvl-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16303260"
/translation="MGKLSKPAVCKRRESPEGDSFAVSAWARKGIEBWRGRC
PGVSGRGLRIAGTGRRELVDRLVRLTLSEEDDRLRLVRLPEKTDGSGDE
KMERVSEPPGSKQLKPEBELQDVSNEDSOGENTFTLIDPDNNGKTRBEITSL
HTLYEVDSVNSPTSSKMLRYKLTVAIDPSKRSVLVNDLDSANPRATKTEPL
DLSSMEKQRAPLRFQDSRLBQSGVHCVENIRRMHYLDLAGIENYISQFGRS
PSVAGSELPPTPSNPTRSRSHPEAIIHPIKRPQGVDPASFFLPTPLAKVSELQR
LRGTODSGHFPVPSKPAQGSVGVHVGARMKPLGPALIPAVSPSAHLAASGAPLIP
SLAPLGHKXKHRAKESQOGCRGLQPLASGSPVLRHRLPALVYVESGAPQVQ
RHHHHHHHHHHHHFYQT"
ORIGIN
Query Match 86.6%; Score 1610; DB 9; Length 1788;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 219 CCAGCATGGGAAATTTCACTCCAAAGCGGCGCGCTGTGCAAGCCGAGGAGACCCCG 278
DB 1 CCAGCATGGGAAATTTCACTCCAAAGCGGCGCGCTGTGCAAGCCGAGGAGACCCCG 60
QY 279 AAGGTGACAGCTTCGCGTGAAGGCTGCTGGGCTGGAAAGGCAATGAGAGTGATCG 338
DB 61 AAGGTGACAGCTTCGCGTGAAGGCTGCTGGGCTGGAAAGGCAATGAGAGTGATCG 120
QY 339 GGAAGCAGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGCTGCGGTTGGCGGACCA 398
DB 121 GGAAGCAGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGCTGCGGTTGGCGGACCA 180
QY 121 GGAAGCAGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGCTGCGGTTGGCGGACCA 180
QY 399 TAGGCGAAGCACCCGGAGCTGTGGGCGACGCTGTGAGAGACAGCTCAGCGAGAG 458
DB 181 TAGGCGAAGCACCCGGAGCTGTGGGCGACGCTGTGAGAGACAGCTCAGCGAGAG 240
QY 459 AGAGAGACGACTTTCCGCTGGAGTGCGCTGCTCTGAGAGACGAGCGGCTGGGCA 518
DB 241 AGAGAGACGACTTTCCGCTGGAGTGCGCTGCTCTGAGAGACGAGCGGCTGGGCA 300
QY 519 GCGGATGAGAGAGATGAGAGATGAGAGACCCCTGCGGCTCCAGAGACGAGC 578
DB 301 GCGGATGAGAGAGATGAGAGATGAGAGACCCCTGCGGCTCCAGAGACGAGC 360
QY 579 TGAAGTTTGAAGAGCTTCAGTGCAGCTGTCCATGAGAGAGACAGCGGCGAGAGTGA 638
DB 361 TGAAGTTTGAAGAGCTTCAGTGCAGCTGTCCATGAGAGAGACAGCGGCGAGAGTGA 420
```

QY 639 CCTTCACTCTGTATGATCTTTGACAACAACGCGAAGTCACTCCGAGAGACATCAACAGCT 698
 DB 421 CCTTCACTCTGTATGATCTTTGACAACAACGCGAAGTCACTCCGAGAGACATCAACAGCT 480
 QY 699 TGGTGCACACCATATGAGGTGGTGGTCTCTGTCAACCATCCGCCAATCCAGCA 758
 DB 481 TGGTGCACACCATATGAGGTGGTGGTCTCTGTCAACCATCCGCCAATCCAGCA 540
 QY 759 AGATGCTGCGGGTAAAGCTCACGTTGGCCCCCGATGGCAGCCAGAGCAAGAGAGCGTCC 818
 DB 541 AGATGCTGCGGGTAAAGCTCACGTTGGCCCCCGATGGCAGCCAGAGCAAGAGAGCGTCC 600
 QY 819 TTGTCAATCAGGCTGACCTGCAAGCGGAGCGCCCGAGCAGAGACCAAGCCCATGAGG 878
 DB 601 TTGTCAATCAGGCTGACCTGCAAGCGGAGCGCCCGAGCAGAGACCAAGCCCATGAGG 660
 QY 879 ACCTGGGAGAGCTGGAGAGAGAGAGAGCGGCTCAGGTTCCAGGGTTCACAGCGGCC 938
 DB 661 ACCTGGGAGAGCTGGAGAGAGAGAGAGCGGCTCAGGTTCCAGGGTTCACAGCGGCC 720
 QY 939 TGGAGCAGTCTGCTGTACCAACCATTTGCTGATGAGAACATGAGAGAGAGAAACGCT 998
 DB 721 TGGAGCAGTCTGCTGTACCAACCATTTGCTGATGAGAACATGAGAGAGAGAAACGCT 780
 QY 999 ACTTAGATCTCGCCGGATGAGAAACTACACGTCCTCAATTTGGGCTTGGCTCCCTTCG 1058
 DB 781 ACTTAGATCTCGCCGGATGAGAAACTACACGTCCTCAATTTGGGCTTGGCTCCCTTCG 840
 QY 1059 TGGCCCAAGAGTCAAGTCTGCCCCCGGACCTCCATCCCATCTGCATCTTGTCCCATG 1118
 DB 841 TGGCCCAAGAGTCAAGTCTGCCCCCGGACCTCCATCCCATCTGCATCTTGTCCCATG 900
 QY 1119 AGCCGGAAGCATATCAGATCCCAACCGAAAGCCCAAGGCGTGAACCGGCTCTCTTCC 1178
 DB 901 AGCCGGAAGCATATCAGATCCCAACCGAAAGCCCAAGGCGTGAACCGGCTCTCTTCC 960
 QY 1179 ACTTCTTGAACATCCCAATGCGCAAGGCTCTGAGATCTTCAGCAACGCGTCCGGGCAACC 1238
 DB 961 ACTTCTTGAACATCCCAATGCGCAAGGCTCTGAGATCTTCAGCAACGCGTCCGGGCAACC 1020
 QY 1239 AGAGCGGAGCAAGCATTTGTGAGTCCCCCAAGGCGCAAGGAGAGAGTGTGGTGG 1298
 DB 1021 AGAGCGGAGCAAGCATTTGTGAGTCCCCCAAGGCGCAAGGAGAGAGTGTGGTGG 1080
 QY 1299 GCCACGTGCGCAGAGGCGCAAGAAACAAGCCCTCTTGGACCGGCAATCCCTGCGGTGT 1358
 DB 1081 GCCACGTGCGCAGAGGCGCAAGAAACAAGCCCTCTTGGACCGGCAATCCCTGCGGTGT 1140
 QY 1359 CCCCCTCCGCGCACTGGCTGCGACGCCGCGCTCTCCCTCCCTAGCCCCCTCGGAC 1418
 DB 1141 CCCCCTCCGCGCACTGGCTGCGACGCCGCGCTCTCCCTCCCTAGCCCCCTCGGAC 1200
 QY 1419 ACAAGAAAGCAACAGCAACCGAGCCAGAGAGCCAGAGGCTGCGGCGCTTCAGAGCAC 1478
 DB 1201 ACAAGAAAGCAACAGCAACCGAGCCAGAGAGCCAGAGGCTGCGGCGCTTCAGAGCAC 1260
 QY 1479 CACTGCGCTCAGGTCGCTCTGCGGCGCGAGACCTTGGGAGAGTGGCCGCTTGG 1538
 DB 1261 CACTGCGCTCAGGTCGCTCTGCGGCGCGAGACCTTGGGAGAGTGGCCGCTTGG 1320
 QY 1539 TGGTGTATGAGAGCGGCGGAGCGCGGCTCAGAGACATGAGACACCAACCAACCATG 1598
 DB 1321 TGGTGTATGAGAGCGGCGGAGCGCGGCTCAGAGACATGAGACACCAACCAACCATG 1380
 QY 1599 AACTATCAACATTAACCACTTTTACAGAGATGAGGCGCTCCCGAGGCGCCAC 1658
 DB 1381 AACTATCAACATTAACCACTTTTACAGAGATGAGGCGCTCCCGAGGCGCCAC 1440
 QY 1659 CTGCGATATGAAGACCCCAACCCCGACACCAAGGACATTAATTATTAATTATG 1718
 DB 1441 CTGCGATATGAAGACCCCAACCCCGACACCAAGGACATTAATTATTAATTATG 1500

QY 1719 TTATATGATGATATATTTATTAATAATTAATTTACTCCATTAATTTAGTACCT 1778
 DB 1501 TTATATGATGATATATTTATTAATAATTAATTTACTCCATTAATTTAGTACCT 1560
 QY 1779 ACATGTAGAAGATCTATGAAACACAGAACTAACTTTTATTTATATGTT 1828
 DB 1561 ACATGTAGAAGATCTATGAAACACAGAACTAACTTTTATTTATATGTT 1610
 RESULT 4
 AF358135 1438 bp mRNA linear PRI 29-MAY-2001
 LOCUS
 DEFINITION Homo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.
 ACCESSION AF358135
 VERSION AF358135.1 GI:14211713
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 AUTHORS Wharton,K.A., Jr., Zimmermann,G., Rouseeet,R. and Scott,M.P.
 TITLE Vertebrate proteins related to Drosophila Naked Cuticle bind
 Dev. Biol. 234 (1), 93-106 (2001)
 JOURNAL
 MEDLINE 2125530
 PUBMED 11356022
 REFERENCE
 AUTHORS Wharton,K.A., Jr., Zimmermann,G. and Scott,M.P.
 TITLE Direct Submission
 Submitted (08-MAR-2001) Pathology and Molecular Biology, University
 of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,
 Dallas, TX 75390-9072, USA
 JOURNAL
 FEATURES
 source
 1..1438
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q12"
 1..1438
 /gene="NKD1"
 8..1420
 /gene="NKD1"
 /note="contains EFX domain; similar to Drosophila
 melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"
 /codon_start=1
 /product="naked cuticle-1"
 /protein_id="AAK57484.1"
 /db_xref="GI:14211714"
 /translation="MGKLSKRAVAACVCKRRESPGDSFVSAAMARKGIEBIMIGRCP
 GGVSPROLACTIGRSTREIVLDYRLDTLSSEEDPRLBYALPREKTDGSGDE
 KKMREVSERPGSKKOLPREILODVMSEDSRQEMFTLLYDNNNGKTRREDITSL
 HTIYEVYDSSVNHSPITSSKRLRYKLTVAIRPDSGSKSVLVMDLDGSRAPRATKPRF
 DLRSWEKQDARLRFQDSKRLSESGTNCVDENIRKRYLDLAGIENTSGPES
 PSVAQSELEPRTPRSNRPRAIINLPHRQGVDFEFLDPLIKVLELQOR
 LRGTODGSKHVAIPKQSKSVGVGVHVARARPKRPGLPAIPVSSAHIAALP
 SLARLGHKKHKAESQSGRGILQPLASGCVLGRHNDRLRYLVVSSQGRPVC
 RHENHNHNHNHNHNHNYQR"
 ORIGIN
 Query Match 77.4%; Score 1438; DB 9; Length 1438;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 217 CCCCAGATGGGGAACCTTCACTCCAAAGCCGCGCGCTGTGCAAGCGCAGAGAGCCC 276
 DB 1 CCCCAGATGGGGAACCTTCACTCCAAAGCCGCGCGCTGTGCAAGCGCAGAGAGCCC 60
 QY 277 GGAAGGTGACAGCTTCCCGCTGAGCGCTGCGGCTCGGAAGGGCATCGAGAGTGGAT 336
 DB 61 GGAAGGTGACAGCTTCCCGCTGAGCGCTGCGGCTCGGAAGGGCATCGAGAGTGGAT 120

QY	333	CGGGAGACAGCCCTGCCCGGGCGGTGTCTCGGGACCCCGACAGCTCGGTTGGCGGGCAC	396
Db	121	CGGGAGACAGCCCTGCCCGGGCGGTGTCTCGGGACCCCGACAGCTCGGTTGGCGGGCAC	180
QY	397	CATPAGGCCAAGCACCCGGGAGCTGTGGGGGACGTTTGTAGAGACAGCTTCAGAGAGA	456
Db	181	CATPAGGCCAAGCACCCGGGAGCTGTGGGGGACGTTTGTAGAGACAGCTTCAGAGAGA	240
QY	457	AGAGGAGGACGACTTTCGGCTGGAAGTGGCCCTGCTCTCTGAGAAAGACTGACGGCTGGG	516
Db	241	AGAGGAGGACGACTTTCGGCTGGAAGTGGCCCTGCTCTCTGAGAAAGACTGACGGCTGGG	300
QY	517	CAGCGGAGATGAGAAAGATGGAAGAGTGAAGCCGACCTTCGCCAGGCTCCAGAGACA	576
Db	301	CAGCGGAGATGAGAAAGATGGAAGAGTGAAGCCGACCTTCGCCAGGCTCCAGAGACA	360
QY	577	GGCGAAGTTTGAAGAGCTCCAGTGGGACGTGTCCATGAGAGAGGACAGCCGGCAGAGAGT	636
Db	361	GGCGAAGTTTGAAGAGCTCCAGTGGGACGTGTCCATGAGAGAGGACAGCCGGCAGAGAGT	420
QY	637	GACCTTCACCCCTGTATGACTTTTGAACAACAACGGGCAAGGTCAACCCAGAGGACATACACG	696
Db	421	GACCTTCACCCCTGTATGACTTTTGAACAACAACGGGCAAGGTCAACCCAGAGGACATACACG	480
QY	697	CTTGTCTGCAACACATCTATGAGTGGTGGACTCTCTGTCAAACACTCCCAACATCCAG	756
Db	481	CTTGTCTGCAACATCTATGAGTGGTGGACTCTCTGTCAAACACTCCCAACATCCAG	540
QY	757	CAAGATGCGCGGGGTAAAGCTCACCGTGCCCCCGCATGGCAGCCAGACCAAGAGAGGT	816
Db	541	CAAGATGCGCGGGGTAAAGCTCACCGTGCCCCCGCATGGCAGCCAGACCAAGAGAGGT	600
QY	817	CCTTGTCAATCAGAGCTGACCTGCAGAGCGCAGAGGCCCGCAGCAGAGACAACAGCCCATCTGA	876
Db	601	CCTTGTCAATCAGAGCTGACCTGCAGAGCGCAGAGGCCCGCAGCAGAGACAACAGCCCATCTGA	660
QY	877	GGAACCTGCGGAGCTGGGAGAAAGACGCGAGCCCGCTCAGGTTCCAGGTTGACAGCCG	936
Db	661	GGAACCTGCGGAGCTGGGAGAAAGACGCGAGCCCGCTCAGGTTCCAGGTTGACAGCCG	720
QY	937	CCTGAGAGCATGTGGCTGCTACCAACATTTGCGCTPATATAGAACATTCGAGAGAGAAACA	996
Db	721	CCTGAGAGCATGTGGCTGCTACCAACATTTGCGCTPATATAGAACATTCGAGAGAGAAACA	780
QY	997	CTACTTATGATCTCGCCGCGGATAGAAAACTACACGTCGCCAATTTGGGCTGTGCCCTTC	1056
Db	781	CTACTTATGATCTCGCCGCGGATAGAAAACTACACGTCGCCAATTTGGGCTGTGCCCTTC	840
QY	1057	CGTGCGCCGAAAGTCGAACTGCCCCCGCGCACCTTCATATCCCACTCGATCTCGTCTCCA	1116
Db	841	CGTGCGCCGAAAGTCGAACTGCCCCCGCGCACCTTCATATCCCACTCGATCTCGTCTCCA	900
QY	1117	TGAGCCCGGAAGCATTCACATCCCAACCGGAAGCCCAAGGCGTGGACCCGGCTCTCTT	1176
Db	901	TGAGCCCGGAAGCATTCACATCCCAACCGGAAGCCCAAGGCGTGGACCCGGCTCTCTT	960
QY	1177	CCACTTCTTTGACACCCCAATCGCAAGGATCTCAGAGCTCCAGCAACGGCTCTGGGAGCAC	1236
Db	961	CCACTTCTTTGACACCCCAATCGCAAGGATCTCAGAGCTCCAGCAACGGCTCTGGGAGCAC	1020
QY	1237	CCAGGACGGGAGCAGCACTTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGT	1296
Db	1021	CCAGGACGGGAGCAGCACTTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGT	1080
QY	1297	GGGCGCACGTGGCCAGAGGGGGCAAGAAACAAGCCCCCTCTGGGACCCCGCATCTCTGCGGT	1356
Db	1081	GGGCGCACGTGGCCAGAGGGGGCAAGAAACAAGCCCCCTCTGGGACCCCGCATCTCTGCGGT	1140
QY	1357	GTCCCCCTCGGCGCACGTGGCTGGCAGAGCCCGGCGCTCTCCCTCGCTAGAGCCCGCTCGG	1416
Db	1141	GTCCCCCTCGGCGCACGTGGCTGGCAGAGCCCGGCGCTCTCCCTCGCTAGAGCCCGCTCGG	1200
QY	1417	GCACAAAGACACAAAGCACCGAGCCAAAGAGAGCAGCAGGCTCTCGGGGCTTCACAGC	1476

Db	1201	GCACGAAGAAGCAACAAGCACCGAGCCGAAGAAAGCCACAGCAGGGCTGGCCGGGGCTGGCAAGC	1260
Qy	1477	ACCACTGGCCCTCAAGGTGGCCCTGTCTCTGGGCGGGAGACCTTGGCGGAGCTGCCGCTT	1536
Db	1261	ACCACTGGCCCTCAAGGTGGCCCTGTCTCTGGGCGGGAGAGACCTGGGGGAGCTGCCGCTT	1320
Qy	1537	GGTGGTGTATGAGAGCCAGGCGGGGCGACCGCGTCCAGAGACATGAGCAACCAACCA	1596
Db	1321	GGTGGTGTATGAGAGCCAGGCGGGGCGACCGCGTCCAGAGACATGAGCAACCAACCA	1380
Qy	1537	TGAACATCACCAACCATTTACCAACCTTCTACTACAGACATPAGAGCCCTCCAGGGGCC	1654
Db	1381	TGAACATCACCAACCATTTACCAACCTTCTACTACAGACATPAGAGCCCTCCAGGGGCC	1438
RESULT 5			
AR405917	AR405917	1438 bp	DNA linear PAT 18-DEC-2003
LOCUS	Sequence 5 from patent US 6630323.		
DEFINITION	AR405917		
ACCESSION	AR405917		
VERSION	AR405917.1	GI:40154965	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1438)		
TITLE	Scott, M., Zeng, W. and Wharton, K.		
JOURNAL	Naked cuticle genes and their uses		
FEATURES	Patent: US 6630323-A 5 07-OCT-2003;		
	Location/Qualifiers		
	1..1438		
ORIGIN	/organism="unknown"		
	/mol_type="genomic DNA"		
Query Match	71.9%;	Score 1336;	DB 6; Length 1438;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 1436;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	217	CCCCGACATGGGGAACTTCACTCCAAAGCGGGCGCGGTGNCAAAGCCAGAGAGAGCC	276
Db	1	CCCCGACATGGGGAACTTCACTCCAAAGCGGGCGCGGTGNCAAAGCCAGAGAGAGCC	60
Qy	277	GGAAGGTGACAGCTTCCGCGCTGAGCGCTGCTCGAGGACTCGAGAGTGGAT	336
Db	61	GGAAGGTGACAGCTTCCGCGCTGAGCGCTGCTCGAGGAGCAATCGAGAGTGGAT	120
Qy	337	CGGAGACAGCGCTGCCCGGGCGGTGTCTCGGAGACCCCGACAGCTGCGGTTGGCGGGCAC	396
Db	121	CGGAGACAGCGCTGCCCGGGCGGTGTCTCGGAGACCCCGACAGCTGCGGTTGGCGGGCAC	180
Qy	397	CATAGGCCGAAGCACCCCGGAGGCTCTGGGGGAGAGTGTTGAGAGACACCGCTCACGAGGA	456
Db	181	CATAGGCCGAAGCACCCCGGAGGCTCTGGGGGAGAGTGTTGAGAGACACCGCTCACGAGGA	240
Qy	457	AGAGAGGACAGCTTTCGGCTGGAAGTGGCCCTGCTCTGAGAGAACTGACCGGGCTGGG	516
Db	241	AGAGAGGACAGCACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAGAACTGACCGGGCTGGG	300
Qy	517	CAGCGGAGATGAGAAAGATGAGAGAGTGAAGCGAACTCTGCCAGGCTTCAAGAAACA	576
Db	301	CAGCGGAGATGAGAAAGATGAGAGAGTGAAGCGAACTCTGCCAGGCTTCAAGAAACA	360
Qy	577	GCTGAAGTTTGAAGAGCTTCCAGTGCAGCTGTCCATGAGAGAGACAGCCCGGCGAGTGG	636
Db	361	GCTGAAGTTTGAAGAGCTTCCAGTGCAGCTGTCCATGAGAGAGACAGCCCGGCGAGTGG	420
Qy	637	GACCTTCAACCTGTATGACTTTTGAACAACAAGGCAAGGTCAACCGAGAGGACATCAACAG	696
Db	421	GACCTTCAACCTGTATGACTTTTGAACAACAAGGCAAGGTCAACCGAGAGGACATCAACAG	480
Qy	697	CTTGCTGACACCAATCTATGAGAGTGTGAGCTCTCTGTCAACCACTCCCAATCCAG	756

Db	481	CTGTGTGACACCAATCTATGAGGTGTGGA	CTCTCTGTCAACCACTCCCAACATCCAG	540
QY	757	CAAGATGCTGGGGGTTAAAGCTCA	CCGTGGCCCCCGATGGCAGCAGCAAGAGCGT	816
Db	541	CAAGATGCTGGGGGTTAAAGCTCA	CCGTGGCCCCCGATGGCAGCAGCAAGAGCGCT	600
QY	817	CCTTGTCANATCAGGCTGACCTTG	CAAGGCGCCGAGGAGAGACCAACCCCATCTGA	876
Db	601	CTTGTCANATCAGGCTGACCTTG	CAAGGCGCCGAGGAGAGACCAACCCCATCTGA	660
QY	877	GGAACCTGCGAGCTGGAGAAAGACA	GCGAGCCCCGCTCAGGTTCCAGGCTGACAGCGG	936
Db	661	GGAACCTGCGAGCTGGAGAAAGACA	GCGAGCCCCGCTCAGGTTCCAGGCTGACAGCGG	720
QY	937	CCTGGAGCAATCTGGCTGCTAC	CAACCATTTGGCTGATGTAGAA	996
Db	721	CCTGGAGCAATCTGGCTGCTAC	CAACCATTTGGCTGATGTAGAA	780
QY	997	CTACTTATGATCTCGCGGGGAT	TGAAAACCTACACGTCGCCAATTTGGGGCTGGGCTCCCTTC	1055
Db	781	CTACTTATGATCTCGCGGGGAT	TGAAAACCTACACGTCGCCAATTTGGGGCTGGGCTCCCTTC	840
QY	1057	CGTGGCCAGAAATCTAGAACTG	CCCCCGCACCTTCCATTCCTCAGTCTCGTCCCA	1114
Db	841	CGTGGCCAGAAATCTAGAACTG	CCCCCGCACCTTCCATTCCTCAGTCTCGTCCCA	900
QY	1117	TGAGCCCGAAGCCATCCCAT	ATCCCAACCCGAAAGCCCAAGGCGTGTGACCCGGGCTCTCTT	1177
Db	901	TGAGCCCGAAGCCATCCCAT	ATCCCAACCCGAAAGCCCAAGGCGTGTGACCCGGGCTCTCTT	960
QY	1177	CCACTTCTCTTGACACCCCAAT	TGCGCAAGGCTCCAGAGCTCCAGCAACGGGCTCCGGGGCAC	1234
Db	961	CCACTTCTCTTGACACCCCAAT	TGCGCAAGGCTCCAGAGCTCCAGCAACGGGCTCCGGGGCAC	1022
QY	1237	CGAAGACGGAGCAAGCACTTT	TGTAGGTCCTCCCAAGGCCCAAGGCGAAGAGTGTGGGTGT	1294
Db	1021	CGAAGACGGAGCAAGCACTTT	TGTAGGTCCTCCCAAGGCCCAAGGCGAAGAGTGTGGGTGT	1080
QY	1297	GGGGCAAGTGGCCAGAGGGGGCA	AGAAACAAGCCCCCTCTGGGACCCCGCATTCCTCGGCGGT	1355
Db	1081	GGGGCAAGTGGCCAGAGGGGGCA	AGAAACAAGCCCCCTCTGGGACCCCGCATTCCTCGGCGGT	1144
QY	1357	GTCCCCCTCCGCGCACTGGG	CTGACAGCCCGGCTCTCTCCCTCCCTAGACCCCTCCGG	1416
Db	1141	GTCCCCCTCCGCGCACTGGG	CTGACAGCCCGGCTCTCTCCCTCCCTAGACCCCTCCGG	1200
QY	1417	GCAACAGAAACAACAAGCA	CCGAGCCAGAGAGCAAGAGGAGCTGCGGGGCTCTCAGGCG	1477
Db	1201	GCAACAGAAACAACAAGCA	CCGAGCCAGAGAGCAAGAGGAGCTGCGGGGCTCTCAGGCG	1260
QY	1477	ACCACTGGGCTCAAGGTGG	CCCTGTCTCTGGGGGGAGAGCACTGGGGGAGCTGGCCGGCTT	1533
Db	1261	ACCACTGGGCTCAAGGTGG	CCCTGTCTCTGGGGGGAGAGCACTGGGGGAGCTGGCCGGCTT	1320
QY	1537	GGTGGTGTATAGAGAC	CGAGCCGGGCAAGCCGGTCCAGAGACATGTAGACACACCAACCA	1596
Db	1321	GGTGGTGTATAGAGAC	CGAGCCGGGCAAGCCGGTCCAGAGACATGTAGACACACCAACCA	1380
QY	1597	TGAAACATACCAACCATTT	ACACACTTTTACAGACATATAGAGCCCTCTCCACAGGGGCC	1654
Db	1381	TGAAACATACCAACCATTT	ACACACTTTTACAGACATATAGAGCCCTCTCCACAGGGGCC	1438

RESULT	6				
LOCUS	CQ721681				
DEFINITION	Sequence 7615 from Patent WO02068579.	1053 bp	DNA		
ACCESSION	CQ721681				linear
VERSION	CQ721681.1	GI:42282538			PAT 03-FEB-2004
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
ORIGIN

1
Venter, C. J., Adams, M. C., Li, P. W. and Myers, E. W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
Patent: WO 02068579-A 7615 06-SEP-2002;
PE Corporation (NY) (US)
Location/Qualifiers
1. .1053
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

Query Match		43.7%; Score 812; DB 6; Length 1053;			
Best Local Similarity		100.0%; Pred. No. 0;			
Matches	812; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	481 AGTGGCCCTGCTCTTGAGAAAGACTGA	CGGGCTGGGCA	CGGAGATGAAAGAAATGGA	540	
DB	81 AGTGGCCCTGCTCTTGAGAAAGCTGA	CGGGCTGGGCA	CGGAGATGAGAAAGATGGA	140	
QY	541 GAGGTAGGGAACCCCTGCCAGGCTCCAAAGACACTGAAGTTTGAAGCTCCAGTG			600	
DB	141 GAGGTAGGGAACCCCTGCCAGGCTCCAAAGAACACTGAAGTTTGAAGGCTCCAGTG			200	
QY	601 CGAGGTGTCCATGAGAGAGGACAGCCGGGACAGATGTGACCTTCAACCTGTATGACTTTGA			660	
DB	201 CGAGGTGTCCATGAGAGAGGACAGCCGGGACAGATGTGACCTTCAACCTGTATGACTTTGA			260	
QY	661 CAACACCGGCAAGGTACACCGGAGAGACATCAACAGCTTGCTGCACACCATCTATGAGGT			720	
DB	261 CAACACCGGCAAGGTACACCGGAGAGACATCAACAGCTTGCTGCACACCATCTATGAGGT			320	
QY	721 GGTGGAATCCCTCTGTCTAACAACCTCCCAACATCCAGCAAGATGTGGGGGTAAAGCTCAC			780	
DB	321 GGTGGAATCTCTCTGTCTAACAACCTCCCAACATCCAGCAAGATGTGGGGGTAAAGCTCAC			380	
QY	781 CGTGGCCCCGATGGGACGACAGACAAAGAGAGCGTCTTGTCAATCAGGCTGACCTTGA			840	
DB	381 CGTGGCCCCGATGGGACGACAGACAAAGAGAGCGTCTTGTCAATCAGGCTGACCTTGA			440	
QY	841 GAGGCGCAAGGCCCCGAGACAGAGACCAACCCCATGAGAGCTTGCGGAGCTTGGAGAAAGA			900	
DB	441 GAGGCGCAAGGCCCCGAGACAGAGACCAACCCCATGAGAGCTTGCGGAGCTTGGAGAAAGA			500	
QY	901 GCAGCGAGCCCGGCTCAGGTTCCAGGGGTGA	CAGCGCGCTGGAGACAGTCTGGCTGTACCA		960	
DB	501 GCAGCGAGCCCGGCTCAGGTTCCAGGGGTGACAGCGCGCTTGAGACAGTCTGGCTGTACCA			560	
QY	961 CCATTGGCTGATGAGAACTCGAGAGAGAAACCACTACTTATGATCTCGCCGGGATAGA			1020	
DB	561 CCATTGGCTGATGAGAACTCGAGAGAGAAACCACTACTTATGATCTCGCCGGGATAGA			620	
QY	1021 AAACCTAACGTCCTCAATTTTGGGCGTGGCTCCCTTCGCTGGCCCAAGAGTCAGAACCTGCC			1080	
DB	621 AAACCTAACGTCCTCAATTTTGGGCGTGGCTCCCTTCGCTGGCCCAAGAGTCAGAACCTGCC			680	
QY	1081 CCCCCGCACTCTCAATCCACTCGATCTCGCTCCATAGACCGGAGACCATTCACATCTCC			1140	
DB	681 CCCCCGCACTCTCAATCCACTCGATCTCGCTCCATAGACCGGAGACCATTCACATCTCC			740	
QY	1141 ACACCGAAAGCCCCAGGCGGTGA	CCCGGCTCTCTTCACTTCTTGAACACCCCAATCGC		1200	
DB	741 ACACCGAAAGCCCCAGGCGGTGA	CCCGGCTCTCTTCACTTCTTGAACACCCCAATCGC		800	
QY	1201 CAAGGTCTCAGAGTCCAGCAAGGCTCCGAGGAGACCCAGAGACGGGAGCAAGCACTTTGT			1260	
DB	801 CAAGGTCTCAGAGTCCAGCAAGGCTCCGAGGAGACCCAGAGACGGGAGCAAGCACTTTGT			860	
QY	1261 GAGGTCCCCCAAGGCCAGGCAAGAGTGTGG			1292	


```

QY 1046 GGCTCCCTTCCGTCGAGGAGTCAAGTCTGCCCCCGGACCTTCAATCCCACTCGA 1105
DB 31966 GGCTCCCTTCCGTCGAGGAGTCAAGTCTGCCCCCGGACCTTCAATCCCACTCGA 32025
QY 1106 TCTGCTCCCATGAGCGGAGGAGTCAATCCCAATCCCAATCCCAATCCCAATCCCA 1165
DB 32026 TCTGCTCCCATGAGCGGAGGAGTCAATCCCAATCCCAATCCCAATCCCAATCCCA 32085
QY 1166 CGGCTCTCTTCCCACTTCTTGAACCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 1225
DB 32086 CGGCTCTCTTCCCACTTCTTGAACCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 32145
QY 1226 CTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1285
DB 32146 CTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 32205
QY 1286 AGTGTGGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1345
DB 32206 AGTGTGGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 32265
QY 1346 ATCCCTGCGGTGTCTCCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1405
DB 32266 ATCCCTGCGGTGTCTCCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 32325
QY 1406 GCGCCCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1465
DB 32326 GCGCCCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 32385
QY 1466 GCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1525
DB 32386 GCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 32445
QY 1526 CTGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1585
DB 32446 CTGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 32505
QY 1586 CACCAACCAATGAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1645
DB 32506 CACCAACCAATGAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 32565
QY 1646 CCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1705
DB 32566 CCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 32625
QY 1706 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1765
DB 32626 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 32685
QY 1766 ATTTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 1825
DB 32686 ATTTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 32745
QY 1826 GTT 1828
DB 32746 GTT 32748

```

```

RESULT 8
AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
AC007608
VERSION AC007608.7 GI:29124038
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169773)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission

```

```

JOURNAL
REFERENCE 2 (bases 1 to 169773)
AUTHORS Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longstre, J., White, S., Tetum, O.,
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
3 (bases 1 to 169773)
REFERENCE DOE Joint Genome Institute.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 169773)
REFERENCE DOE Joint Genome Institute.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
5 (bases 1 to 169773)
REFERENCE DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 19, 2003 this sequence version replaced gi:18129388.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
1..169773
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-401P9"
ORIGIN
Query Match 42.1%; Score 783; DB 9; Length 169773;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGCTCCCTTCCGTCGAGGAGTCAAGTCTGCCCCCGGACCTTCAATCCCACTCGA 1105
DB 78524 GGCTCCCTTCCGTCGAGGAGTCAAGTCTGCCCCCGGACCTTCAATCCCACTCGA 78465
QY 1106 TCTGCTCCCATGAGCGGAGGAGTCAATCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 1165
DB 78464 TCTGCTCCCATGAGCGGAGGAGTCAATCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 78405
QY 1166 CGGCTCTCTTCCCACTTCTTGAACCCCAATCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 1225
DB 78404 CGGCTCTCTTCCCACTTCTTGAACCCCAATCCCAATCCCAATCCCAATCCCAATCCCAATCCCA 78345
QY 1226 CTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1285
DB 78344 CTCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 78285
QY 1286 AGTGTGGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1345
DB 78284 AGTGTGGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 78225
QY 1346 ATCCCTGCGGTGTCTCCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1405
DB 78224 ATCCCTGCGGTGTCTCCCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 78165
QY 1406 GCGCCCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1465

```

Db	78164	GCCCCCTCGGGGCACAAGAAAGCAAGAGACCCAGCCAGAGAGAGCCAGAGGGCTGCGG	78105
Qy	1466	GGCCTGCAGGCAACACTGGGCTCAGGTGGCCCTGTCTGGGGCGGAGACCTTGCGGAG	1525
Db	78104	GGCTGAGGACAACACTGGGCTCAGGTGGCCCTGTCTGGGGGCGGAGACCTTGCGGAG	78045
Qy	1526	CTGGCCGCGCTTGGTGGTGTATGAGAGCCGAGGAGCCGGAGCCGGTCCAGAGACATGAGCAC	1585
Db	78044	CTGGCCGCGCTTGGTGGTGTATGAGAGCCGAGGAGCCGGTCCAGAGACATGAGCAC	77985
Qy	1586	CACCAACACATGAAACATCACCAACCTTACCACTTCTTACACAGCATAGAGCCCTCC	1645
Db	77984	CACCAACACATGAAACATCACCAACCTTACCACTTCTTACACAGCATAGAGCCCTCC	77925
Qy	1646	CGAGGGCCCCACCTGCGCATATAGAGAGCCCAACCCCGACACCAAGAGCATTTATTT	1705
Db	77924	CGAGGGCCCCACCTGCGCATATAGAGAGCCCAACCCCGACACCAAGAGCATTTATTT	77865
Qy	1706	CTATTAAATTGTTATTTATGATGATTTATTTATTAATTAATTTGTTACTCCATAAT	1765
Db	77864	CTATTAAATTGTTATTTATGATGATTTATTTATTAATTAATTTGTTACTCCATAAT	77805
Qy	1766	ATTAGCTAGCCCTACATGTAAGAAAGATCTATGAAACAAGAACTAACTTTATTTATAT	1825
Db	77804	ATTAGCTAGCCCTACATGTAAGAAAGATCTATGAAACAAGAACTAACTTTATTTATAT	77745
Qy	1826	GTT 1828	
Db	77744	GTT 77742	

[illegible]

```

National Laboratory
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1..170791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-147B17"

```

Query Match		42.1%;	Score 783;	DB 9;	Length 170791;	
Best Local Similarity		100.0%;	Pred. No. 0;			
Matches 783;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1046	GGCTCCCTTCCTCGTGAGCCCGAGAACTCGAATCTGAGCTCCCTCCCGCACCTCTCAATCCCACTCGA				1105
DB	28069	GGCTCCCTTCCTCGTGAGCCCGAGAACTCGAATCTGAGCTCCCTCCCGCACCTCTCAATCCCACTCGA				2801
QY	1106	TCCTGGCTCCCATGAGCCGGGAAGCCATCCATCTCCACCCGAAAGCCCCAAAGCGGTGAC				1165
DB	28009	TCCTGGCTCCCATGAGCCGGGAAGCCATCCATCTCCACCCGAAAGCCCCAAAGCGGTGAC				2795
QY	1166	CCGGCTCTCTTCACCTTCTCTTGAACCCCAATGCGCAAGGCTTCAGAGCTTCAGCAACGG				1225
DB	27949	CCGGCTCTCTTCACCTTCTCTTGAACCCCAATGCGCAAGGCTTCAGAGCTTCAGCAACGG				2789
QY	1226	CTCCGGGGGCAACCCAGAACGGGAAGCAACCTTTGTAGGTCTCCCGCAAGGCCCAAGGGCAAG				1285
DB	27889	CTCCGGGGGCAACCCAGAACGGGAAGCAACCTTTGTAGGTCTCCCGCAAGGCCCAAGGGCAAG				2783
QY	1286	AGTGTGGGTGTGGGCGCACAGTGCGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGACCCGGCC				1345
DB	27829	AGTGTGGGTGTGGGCGCACAGTGCGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGACCCGGCC				2777
QY	1346	ATCCCTGCGGTGTCTCCCTTCGCGCCACCTGGCTGCGAGCCCGGCCCTCTCTCCCTCCTTA				1405
DB	27769	ATCCCTGCGGTGTCTCCCTTCGCGCCACCTGGCTGCGAGCCCGGCCCTCTCTCCCTCCTTA				2771
QY	1406	GGCCCCCTCGGGGCAAGAAAGCAACAAGACCGGAGCCAAAGAGAGCCAGAGGGCTGCGCG				1465
DB	27709	GGCCCCCTCGGGGCAAGAAAGCAACAAGACCGGAGCCAAAGAGAGCCAGAGGGCTGCGCG				2765
QY	1466	GGCCTGACAGGACCACTGGCCTCAAGTGGCCCTGTCTGGGGCGGAGACCTGCGGAG				1529
DB	27649	GGCCTGACAGGACCACTGGCCTCAAGTGGCCCTGTCTGGGGCGGAGACCTGCGGAG				2759
QY	1526	CTGCCCCGCTTGTGTGTGTATGAAAGCGACGGCCGGGACCGCGTCCAGAGCATGAGAC				1585
DB	27589	CTGCCCCGCTTGTGTGTGTATGAAAGCGACGGCCGGGACCGCGTCCAGAGCATGAGAC				2753
QY	1586	CACCAACCAATGAAACATCACCACTATTACACCACTTCTACACGAGATAGAGCCCTCC				1645
DB	27529	CACCAACCAATGAAACATCACCACTATTACACCACTTCTACACGAGATAGAGCCCTCC				2747
QY	1646	CCAGGGCCCCACCTTCCCATATGAAGACCCCAACCCCGAGACCAACAAGGCAATTATAT				1705
DB	27469	CCAGGGCCCCACCTTCCCATATGAAGACCCCAACCCCGAGACCAACAAGGCAATTATAT				2741
QY	1706	CTATTATTAATATGTATATGATATGATATATATGTTATTAATATATATGTTACTCCACTAT				1765
DB	27409	CTATTATTAATATGTATATGATATGATATATATGTTATTAATATATATGTTACTCCACTAT				2735
QY	1766	ATTTAGCTAGCGCTACATGTAGAAAGATCTATGAAACAAGAACTTAACTTTATATAT				1825
DB	27349	ATTTAGCTAGCGCTACATGTAGAAAGATCTATGAAACAAGAACTTAACTTTATATAT				2729
QY	1826	GTT 1828				
DB	27289	GTT 27287				

```

RESULT 10
AC145238
LOCUS
DEFINITION Pan troglodytes clone CH251-568P19, WORKING DRAFT SEQUENCE, 25
ordered pieces.
AC145238
VERSION
AC145238.2 GI:33667125
KEYWORDS HTG, HTGS_PHASE2, HTGS_DRAFT.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 178790)
Antonielli, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Caraga, K.,
Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P.,
Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Latic, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Mastello, C.,
Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E.,
Prasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,
Sison, C., Stancirpop, S., Thomas, J.W., Thomas, P.J., Teipouri, V.,
Vogt, J.L., Wetherby, K.D., Young, A. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 178790)
Green, E.D.
Direct Submission
Submitted (18-JUN-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gathersburg, MD 20877, USA
3 (bases 1 to 178790)
Green, E.D.
Direct Submission
Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gathersburg, MD 20877, USA
On Aug 15, 2003 this sequence version replaced gi:31880081.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@hgri.nih.gov
----- Project Information
Center project name: esk
Center clone name: 568P19

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

```

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171494 bases at least Q40
Consensus quality: 174120 bases at least Q30
Consensus quality: 175495 bases at least Q20
Insert size: 149000; agarose-efp
Quality coverage: 176390; sum-of-contigs
Quality coverage: 10.92x in Q20 bases; agarose-efp
Quality coverage: 9.22x in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces

```

FEATURES
    source
        1..178790
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
            /db_xref="taxon:9598"
            /clone="CH251-568P19"
            /clone_1fb="CH251"
            /clone_1lb="CH251"
            /note="clon overlaps with GenBank Accession Number AC145175 clone CH251-32B1 (center project name esj)"
            1..2604
                /note="assembly_fragment"
                clone_end:576
                vector_side:left"
            2705..7790
                /note="assembly_fragment"
                7891..12833
                    /note="assembly_fragment"
                    163753
                        Location/Qualifiers
                            1..178790
                                * is believed to be correct as given, however the sizes
                                * of the gaps between them are based on estimates that have
                                * provided by the submitter.
                                * This sequence will be replaced
                                * by the finished sequence as soon as it is available and
                                * the accession number will be preserved.
                                1
                                    2604: contig of 2604 bp in length
                                2605
                                    2704: gap of unknown length
                                2705
                                    7790: contig of 5086 bp in length
                                7791
                                    7890: gap of unknown length
                                7891
                                    12833: contig of 4943 bp in length
                                12834
                                    12933: gap of unknown length
                                12934
                                    14664: contig of 1731 bp in length
                                14665
                                    14764: gap of unknown length
                                14765
                                    32237: contig of 17473 bp in length
                                32238
                                    32337: gap of unknown length
                                32338
                                    35356: contig of 3019 bp in length
                                35357
                                    35456: gap of unknown length
                                35457
                                    39292: contig of 3836 bp in length
                                39293
                                    39392: gap of unknown length
                                39393
                                    41210: contig of 1818 bp in length
                                41211
                                    43762: contig of 2452 bp in length
                                43763
                                    43863: gap of unknown length
                                43864
                                    46766: contig of 2904 bp in length
                                46767
                                    46866: gap of unknown length
                                46867
                                    53405: contig of 6539 bp in length
                                53406
                                    53505: gap of unknown length
                                53506
                                    55334: contig of 1829 bp in length
                                55335
                                    55434: gap of unknown length
                                55435
                                    69095: contig of 13661 bp in length
                                69096
                                    69195: gap of unknown length
                                69196
                                    78811: contig of 9616 bp in length
                                78812
                                    78911: gap of unknown length
                                78912
                                    89715: contig of 10804 bp in length
                                89716
                                    89815: gap of unknown length
                                89816
                                    97259: contig of 7444 bp in length
                                97260
                                    97359: gap of unknown length
                                97360
                                    99238: contig of 1879 bp in length
                                97360
                                    99238: gap of unknown length
                                99239
                                    99338: gap of unknown length
                                99339
                                    101324: contig of 1986 bp in length
                                101325
                                    101424: gap of unknown length
                                101425
                                    112068: contig of 10664 bp in length
                                112069
                                    112168: gap of unknown length
                                112169
                                    118779: contig of 6511 bp in length
                                118780
                                    118779: gap of unknown length
                                118780
                                    120680: contig of 1901 bp in length
                                120681
                                    120780: gap of unknown length
                                120781
                                    135732: contig of 14952 bp in length
                                135733
                                    135832: gap of unknown length
                                135833
                                    138723: contig of 2891 bp in length
                                138724
                                    138823: gap of unknown length
                                138824
                                    163652: contig of 24829 bp in length
                                163653
                                    163752: gap of unknown length
                                163753
                                    178790: contig of 15038 bp in length.
                                Location/Qualifiers
                                    1..178790
                                        /organism="Pan troglodytes"
                                        /mol_type="genomic DNA"
                                        /db_xref="taxon:9598"
                                        /clone="CH251-568P19"
                                        /clone_1fb="CH251"
                                        /clone_1lb="CH251"
                                        /note="clon overlaps with GenBank Accession Number AC145175 clone CH251-32B1 (center project name esj)"
                                        1..2604
                                            /note="assembly_fragment"
                                            clone_end:576
                                            vector_side:left"
                                        2705..7790
                                            /note="assembly_fragment"
                                            7891..12833
                                                /note="assembly_fragment"
                                                163753
                                                    Location/Qualifiers
                                                        1..178790
                                                            * is believed to be correct as given, however the sizes
                                                            * of the gaps between them are based on estimates that have
                                                            * provided by the submitter.
                                                            * This sequence will be replaced
                                                            * by the finished sequence as soon as it is available and
                                                            * the accession number will be preserved.
                                                            1
                                                                2604: contig of 2604 bp in length
                                                            2605
                                                                2704: gap of unknown length
                                                            2705
                                                                7790: contig of 5086 bp in length
                                                            7791
                                                                7890: gap of unknown length
                                                            7891
                                                                12833: contig of 4943 bp in length
                                                            12834
                                                                12933: gap of unknown length
                                                            12934
                                                                14664: contig of 1731 bp in length
                                                            14665
                                                                14764: gap of unknown length
                                                            14765
                                                                32237: contig of 17473 bp in length
                                                            32238
                                                                32337: gap of unknown length
                                                            32338
                                                                35356: contig of 3019 bp in length
                                                            35357
                                                                35456: gap of unknown length
                                                            35457
                                                                39292: contig of 3836 bp in length
                                                            39293
                                                                39392: gap of unknown length
                                                            39393
                                                                41210: contig of 1818 bp in length
                                                            41211
                                                                43762: contig of 2452 bp in length
                                                            43763
                                                                43863: gap of unknown length
                                                            43864
                                                                46766: contig of 2904 bp in length
                                                            46767
                                                                46866: gap of unknown length
                                                            46867
                                                                53405: contig of 6539 bp in length
                                                            53406
                                                                53505: gap of unknown length
                                                            53506
                                                                55334: contig of 1829 bp in length
                                                            55335
                                                                55434: gap of unknown length
                                                            55435
                                                                69095: contig of 13661 bp in length
                                                            69096
                                                                69195: gap of unknown length
                                                            69196
                                                                78811: contig of 9616 bp in length
                                                            78812
                                                                78911: gap of unknown length
                                                            78912
                                                                89715: contig of 10804 bp in length
                                                            89716
                                                                89815: gap of unknown length
                                                            89816
                                                                97259: contig of 7444 bp in length
                                                            97260
                                                                97359: gap of unknown length
                                                            97360
                                                                99238: contig of 1879 bp in length
                                                            97360
                                                                99238: gap of unknown length
                                                            99239
                                                                99338: gap of unknown length
                                                            99339
                                                                101324: contig of 1986 bp in length
                                                            101325
                                                                101424: gap of unknown length
                                                            101425
                                                                112068: contig of 10664 bp in length
                                                            112069
                                                                112168: gap of unknown length
                                                            112169
                                                                118779: contig of 6511 bp in length
                                                            118780
                                                                118779: gap of unknown length
                                                            118780
                                                                120680: contig of 1901 bp in length
                                                            120681
                                                                120780: gap of unknown length
                                                            120781
                                                                135732: contig of 14952 bp in length
                                                            135733
                                                                135832: gap of unknown length
                                                            135833
                                                                138723: contig of 2891 bp in length
                                                            138724
                                                                138823: gap of unknown length
                                                            138824
                                                                163652: contig of 24829 bp in length
                                                            163653
                                                                163752: gap of unknown length
                                                            163753
                                                                178790: contig of 15038 bp in length.

```


Qy	1646	CCAGGGCCCCACCCCGCATATGAGAGACCCACCCCGACACCAAGGCATTATTATT	1705
Db	181067	CCAGGGCCCCACCCCGCATATGAGAGACCCACCCCGACACCAAGGCATTATTATT	181126
Qy	1706	CTATTAAATTATGTTATTATGATGATTTATTTGTTATTAATTATTTAGTTACTCACTAAT	1765
Db	181127	CTATTAAATTATGTTATTATGATGATTTATTTGTTATTAATTATTTAGTTACTCACTAAT	181186
Qy	1766	ATTTAGCTAGCCTTACATGTGAAAGATCTATGGAAACACAGAACTTAACCTTTATTTATAT	1825
Db	181187	ATTTAGCTAGCCTTACATGTGAAAGATCTATGGAAACACAGAACTTAACCTTTATTTATAT	181246
Qy	1826	GTT 1828	
Db	181247	GTT 181249	
RESULT 12			
LOCUS	AX557538	1950 bp	DNA
DEFINITION	Sequence 126 from Patent W00248370.	linear	PAT 27-NOV-2002
VERSION	AX557538		
KEYWORDS	AX557538.1 GI:25900445		
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
JOURNAL	1		
FEATURES	Sun, Y., Reippon, H., Ghosh, M. G. and Liu, C.		
SOURCE	Compositions and methods relating to colon specific genes and		
	Diadexus, Inc. (US)		
	Patent: WO 0248370-A 126 20-JUN-2002;		
	Location/Qualifiers		
	1..1950		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
	Query Match	7.4%;	Score 137; DB 6; Length 1950;
	Best Local Similarity	100.0%;	Pred. No. 8.3e-64;
	Matches 137; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	281	GGTGACAGCTTCGCGTGAGCCCTGCGTGGAGGGGACATCGAGAGTGGATCGGG	340
Db	838	GGTGACAGCTTCGCGTGAGCCCTGCGTGGAGGGGACATCGAGAGTGGATCGGG	897
Qy	341	AGACAGCGGTGCCCCGGCGGTGTCTCGGAACCCCGACAGCTGCGGTTGGCGGGACACATA	400
Db	898	AGACAGCGGTGCCCCGGCGGTGTCTCGGAACCCCGACAGCTGCGGTTGGCGGGACACATA	957
Qy	401	GGCCGAAGACCCGGGA 417	
Db	958	GGCCGAAGACCCGGGA 974	
RESULT 13			
LOCUS	CQ741468	2340 bp	DNA
DEFINITION	Sequence 27402 from Patent W002068579.	linear	PAT 03-FEB-2004
VERSION	CQ741468		
KEYWORDS	CQ741468.1 GI:42349559		
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
	1		
	Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.		
	Kites, such as nucleic acid arrays, comprising a majority of		
	humanexons or transcripts, for detecting expression and other uses		

```

JOURNAL      Patent: WO 02068579-A 27402 06-SEP-2002;
FEATURES      Location (NY) (US)
SOURCE        Location/Qualifiers
              1. 2340
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"

ORIGIN
Query Match      7.3%; Score 136; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 3e-63;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      280 AGGTGAACAGCTTGCCCTGAGCCCTGCTGCTGCTCGGAAGGCATCGAAGAGTGCATCG 339
DB      1524 AGGTGAACAGCTTGCCCTGAGCCCTGCTGCTGCTCGGAAGGCATCGAAGAGTGCATCG 1583
OY      340 GAGACACGCGTGCCTGGCGGGCGGTGTCTCGGGAGCCCGACAGCTCGGTTCGCGGACCAT 399
DB      1584 GAGACACGCGTGCCTGGCGGGCGGTGTCTCGGGAGCCCGACAGCTCGGTTCGCGGACCAT 1643
OY      400 AGGCCGAAGCACCCTGG 415
DB      1644 AGGCCGAAGCACCCTGG 1659

RESULT 14
AC145003      194869 bp      DNA      linear      HTG 03-JUL-2003
LOCUS      AC145003
DEFINITION      Papio anubis clone RP41-205A15, WORKING DRAFT SEQUENCE, 9 ordered
VERSION      AC145003
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE      Papio anubis (olive baboon)
ORGANISM      Papio anubis
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
              Cercopithecinae; Papio.
              1 (bases 1 to 194869)
REFERENCE      Blakeley, R.W., Bouffard G.G., Binkley, C., Brooks, S., Carriaga, K.,
AUTHORS      Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
              Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-U., Hu, P.,
              Hurtle, B., Idol, J.R., Karlins, E., Kwong, P., Latic, P., Lee-Lin, S.-Q.,
              Legaas, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masello, C.,
              Mackeri, B., McDowell, J., Padirigan, C., Pearson, R., Portnoy, M.E.,
              Prasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,
              Sison, C., Stantirlop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
              Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.
              NISC Comparative Sequencing Initiative
              Unpublished
              2 (bases 1 to 194869)
REFERENCE      Green, E.D.
AUTHORS      Direct Submission
TITLE      Direct Submission
JOURNAL      Submitted (30-MAY-2003) NIH Intramural Sequencing Center, 8717
              Groveomont Circle, Gathersburg, MD 20877, USA
              3 (bases 1 to 194869)
REFERENCE      Green, E.D.
AUTHORS      Direct Submission
TITLE      Direct Submission
JOURNAL      Submitted (03-JUL-2003) NIH Intramural Sequencing Center, 8717
              Groveomont Circle, Gathersburg, MD 20877, USA
              On Jul 3, 2003 this sequence version replaced gi:31193878.
COMMENT      ----- Genome Center
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc.zoo@ngri.nih.gov
              ----- Project Information
              Center project name: dtc
              Center clone name: 205A15

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 193209 bases at least Q40
 Consensus quality: 193760 bases at least Q30
 Consensus quality: 193984 bases at least Q20
 Insert size: 180000; agarose-fp
 Insert size: 194069; sum-of-contigs
 Quality coverage: 14.34x in Q20 bases; agarose-fp
 Quality coverage: 13.30x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

```

1      7609      7608: contig of 7608 bp in length
*      7709      7708: gap of unknown length
*      10152     10251: contig of 2443 bp in length
*      10252     44876: gap of unknown length
*      44877     44976: contig of 34625 bp in length
*      44977     76344: gap of unknown length
*      76345     76444: contig of 31368 bp in length
*      76445     78547: gap of unknown length
*      78548     78547: contig of 2103 bp in length
*      78549     78647: gap of unknown length
*      145819    145819: contig of 67172 bp in length
*      145820    145919: gap of unknown length
*      145920    163253: contig of 17334 bp in length
*      163254    163353: gap of unknown length
*      163354    166218: contig of 2765 bp in length
*      166219    166218: gap of unknown length
*      166219    194869: contig of 28651 bp in length.

```

FEATURES

```

source
  1. 194869
    /location/Qualifiers
      1. 194869
        /organism="Papio anubis"
        /mol_type="genomic DNA"
        /db_xref="taxon:9555"
        /clone="RP41-205A15"
        /clone_lib="RP41"
        1. 31748
          /note="clone overlaps with GenBank Accession Number
          AC145043 clone RP41-2B13 (center project name dtg)"
        1. 7608
          /note="assembly_fragment"
          clone_end:17
          vector_side:left
          7709..10151
            /note="assembly_fragment"
            10252..44876
              /note="assembly_fragment"
              44977..76344
                /note="assembly_fragment"
                76445..78547
                  /note="assembly_fragment"
                  78648..145819
                    /note="assembly_fragment"
                    misc_feature
                      145920..163253
                        /note="assembly_fragment"
                        misc_feature
                          163354..166218
                            /note="assembly_fragment"
                            misc_feature
                              166219..194869
                                /note="assembly_fragment"
                                misc_feature
                                  194870..194869
                                    /note="assembly_fragment"

```

ORIGIN

```

Query Match      6 7%; Score 124; DB 2; Length 194869;
Best Local Similarity 99.4%; Pred. No. 1.7e-56;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
misc_feature      145920..163253
                  /note="assembly_fragment"
misc_feature      163354..166218
                  /note="assembly_fragment"
misc_feature      166219..194869
                  /note="assembly_fragment"
                  clone_end:176
                  vector_side:right

```

```

Qy      1171 CTCCTTCACCTTCCTTGACACCCCAATCCGCAAGTCTCAGAGTCCAGACGCTCCG 1230
Db      16651 CTCCTTCACCTTCCTTGACACCCCAATCCGCAAGTCTCAGAGTCCAGACGCTCCG 16710
Qy      1231 GGGCACCACGACGCGGAGCAAGCACTTGTGAGGTCCCAAGGCCGAGGCAAGAGTGT 1290
Db      16711 GGGCACCACGACGCGGAGCAAGCACTTGTGAGGTCCCAAGGCCGAGGCAAGAGTGT 16770
Qy      1291 GGGGTGCGGCGCAGTGGCGGCAAGGCGGCAAGAACAGCCCTCTGAGACCCGCC 1345
Db      16771 GGGGTGCGGCGCAGTGGCGGCAAGGCGGCAAGAACAGCCCTCTGAGACCCGCC 16825

```

RESULT 15

```

AC145043
LOCUS      1948189 bp DNA linear HTG 26-JUN-2003
DEFINITION Papio anubis clone RP41-2B13, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
AC145043
AC145043.2 GI:32261339
VERSION    HTG: HTGS PHASR2; HTGS DRAFT.
KEYWORDS   Papio anubis (olive baboon)
SOURCE     Papio anubis
ORGANISM   Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
            Cercopitheciinae; Papio.
            1 (bases 1 to 1948189)

```

REFERENCE

```

AUTHORS    Antonellis,A., Ayale,K., Beckstrom-Stenberg,S.M., Benjamin,B.,
            Blakesley,R.W., Boufford,G.G., Brinkley,C., Brooks,S., Carliaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Ihn,S.-O.,
            Legaspi,R., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schneider,M.G., Shah,K.,
            Sison,C., Stantridop,S., Thomas,J.W., Thomas,P.J., Tsipouris,V.,
            Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
            NISC Comparative Sequencing Initiative
            Unpublished
            2 (bases 1 to 1948189)
            Green,E.D.
            Direct Submission
            Submitted (04-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaitheersburg, MD 20877, USA
            3 (bases 1 to 1948189)
            Green,E.D.
            Direct Submission
            Submitted (26-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaitheersburg, MD 20877, USA
            On Jun 26, 2003 this sequence version replaced gi:31376430.

```

REFERENCE

```

AUTHORS    Green,E.D.
TITLE      Direct Submission
JOURNAL    Submitted (04-JUN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaitheersburg, MD 20877, USA
            On Jun 26, 2003 this sequence version replaced gi:31376430.
COMMENT    ----- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc.zoo@nih.gov
            Project Information
            Center project name: dtg
            Center clone name: 002B13

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 16:38:42 ; Search time 1269 Seconds
(without alignments)
7690.050 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859

Sequence: 1 gaattcgccctctactaacg.....aaaaaaaaaagcgccgc 1859

Scoring table: ~~OLIGOBLAST~~
Gapop 60.0 , Gapext 60.0

Searched: 413486 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04.*

1: geneseqn19808.*
2: geneseqn19908.*
3: geneseqn20008.*
4: geneseqn2001a8.*
5: geneseqn2001b8.*
6: geneseqn2002a8.*
7: geneseqn2002b8.*
8: geneseqn2003a8.*
9: geneseqn2003b8.*
10: geneseqn2003c8.*
11: geneseqn2003d8.*
12: geneseqn2004a8.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1859	100.0	1859	10	AAL51511 Human Nkd
2	1336	71.9	1438	3	AAAG3925 DNA encod
3	783	42.1	2379	4	AAL03214 Human rep
4	783	42.1	2379	4	AAL03216 Human rep
5	590	31.7	590	10	AAL51525 Human Nkd
6	405	21.8	1561	5	AAAG2082 DNA encod
7	377	20.3	427	5	ACH21474 Human adu
8	218	11.7	1743	5	AAAG78752 DNA encod
9	218	11.7	1743	5	AAAG2081 DNA encod
10	218	11.7	1743	10	ADCG1966 Human nov
11	215	11.6	215	10	AAL51526 Human Nkd
12	148	8.0	148	10	AAL51522 Human Nkd
13	137	7.4	598	12	ACH74795 Human gen
14	137	7.4	1950	6	ABO73820 Human col
15	134	7.2	134	10	AAL51518 Human Nkd
16	133	7.2	133	12	ACH88495 Human gen
17	128	6.9	128	10	AAL51524 Human Nkd
18	120	6.5	120	10	AAL51514 Human Nkd
19	108	5.8	639	8	ACD05637 cDNA enco
20	107	5.8	107	10	AAL51520 Human Nkd
21	96	5.2	96	10	AAL51521 Human Nkd

22	88	4.7	554	12	ACH74420 Human gen
23	86	4.6	179	12	ACH88120 Human gen
24	85	4.6	85	10	AAL51523 Human Nkd
25	67	3.6	67	10	AAL51519 Human Nkd
26	58	3.1	670	6	ABV73487 Human cyt
27	58	3.1	1264	12	ADO24765 Dog TRAIL
28	57	3.1	417	6	ABV73496 Human cyt
29	57	3.1	510	6	ABV73485 Human cyt
30	56	3.0	636	6	ABV73493 Human cyt
31	56	3.0	651	6	ABV73476 Human cyt
32	56	3.0	686	6	ABV73486 Human cyt
33	55	3.0	707	6	ABX04893 Conus sp
34	55	3.0	547	6	ABV73478 Human cyt
35	55	3.0	1245	6	ABV73473 Human cyt
36	54	2.9	1874	10	ADG82682 Terpenoid
37	54	2.9	1874	10	ADG82678 Terpenoid
38	54	2.9	2605	10	ADG82704 Terpenoid
39	53	2.9	1434	6	ABA92108 Soybean p
40	53	2.9	2755	10	ADG86813 Pecunia p
41	53	2.9	3281	10	ADG32558 Murtine tr
42	52	2.8	577	6	ABV73482 Human cyt
43	52	2.8	1124	6	AAL46702 B tropica
44	52	2.8	1139	6	AAL46704 B tropica
45	52	2.8	1168	6	AAL46703 B tropica

ALIGNMENTS

RESULT 1	AAL51511	standard, DNA, 1859 BP.
ID	AAL51511	
AC	AAL51511;	
DT	24-APR-2003	(first entry)
DE	Human Nkd (hNkd)	protein coding sequence.
KW	Human; gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;	
KW	colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	CDS	224..1636
FT		/*tag= a
FT		/product= "Human Nkd protein"
PN	WO200292832-A2.	
XX	21-NOV-2002.	
XX	27-NOV-2001; 2001WO-US044092.	
XX	27-NOV-2000; 2000US-0252884P.	
XX	16-MAY-2001; 2001US-0291109P.	
XX	01-OCT-2001; 2001US-0325571P.	
XX	(CHIR) CHIRON CORP.	
XX	Rohan M, Chan V, Yan D,	
XX	WPI; 2003-129303/12.	
XX	P-PSDB; AAO16389.	
PT	New human and non-human primate homologues of Nkd protein, and Nkd genes,	
PT	useful for treating cancer involving aberrant Wnt signalling, e.g. colon	
PT	cancer, head and neck cancer, ovarian cancer, or breast cancer.	
PS	Claim 4; Fig 3; 99gp; English.	
CC	The invention comprises the amino acid and coding sequence of the human	

CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
CC protein of the invention
XX

SQ Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 Other;

Query Match 100.0%; Score 1859; DB 10; Length 1859;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATTCGCGCTTCAATACGCTCATATAGGGGAGAGAGTGTAAACAAGSAGATACG 60
Db 1 GAATTCGCGCTTCAATACGCTCATATAGGGGAGAGAGTGTAAACAAGSAGATACG 60
QY 61 CGGGAGATCGGGCCGCGCGACGCGGCGAGAGAGCGCTCCGCGCGCGCTCGGGCTCCG 120
Db 61 CGGGAGATCGGGCCGCGCGCGACGCGGCGAGAGAGCGCTCCGCGCGCGCTCGGGCTCCG 120
QY 121 CTCGGCTCGGGGGCTGCTTGGGAGAGAGAGACCAAGAGGCGCCAGCGCCGCGGCGC 180
Db 121 CTCGGCTCGGGGGCTGCTTGGGAGAGAGAGACCAAGAGGCGCGCCAGCGCGCCGCGGCGC 180
QY 181 GGGGCGATGGCTTAGGGAGCGCTCCGGCGCGCGAGCGCCAGCATGGGGAAACCTTCACTC 240
Db 181 GGGGCGATGGCTTAGGGAGCGCTCCGGCGCGCGAGCGCCAGCATGGGGAAACCTTCACTC 240
QY 241 CAAGCCGCGCGCGCTGTGTGCAAGCGCAGAGAGAGCCGGAAGGTGACAGCTTCCGCGTAG 300
Db 241 CAAGCCGCGCGCGCTGTGTGCAAGCGCAGAGAGAGCCGGAAGGTGACAGCTTCCGCGTAG 300
QY 301 CGCTGCTCGGGCTGGAAGGCGCATTCAGAGAGTGAATGGGAGACAGCGCTCCGCGCGCG 360
Db 301 CGCTGCTCGGGCTGGAAGGCGCATTCAGAGAGTGAATGGGAGACAGCGCTCCGCGCGCG 360
QY 361 TGTCTCGGGAGACCCGACAGCTGCGGTTGGCGGGACCATAGCGCGAAGCAACCGGAGACT 420
Db 361 TGTCTCGGGAGACCCGACAGCTGCGGTTGGCGGGACCATAGCGCGGAGCAACCGGAGACT 420
QY 421 CGTGGGCGACGTTGTGAAGACACGCTCAAGGAGAAAGAGAGACGACTTTCGGCTTGA 480
Db 421 CGTGGGCGACGTTGTGAAGACACGCTCAAGGAGAAAGAGAGACGACTTTCGGCTTGA 480
QY 481 AGTGGCCCTGCTCTGTAAGAAAGTGAAGGCTGGGCGAGGAGATGAGAAAGATGGA 540
Db 481 AGTGGCCCTGCTCTGTAAGAAAGTGAAGGCTGGGCGAGGAGATGAGAAAGATGGA 540
QY 541 GAGAGTGAAGCAACCTGCGCAGGCTCAAGAGACAGCTGAAGTTTGAAGAGCTTCAAGT 600
Db 541 GAGAGTGAAGCAACCTGCGCAGGCTCAAGAGACAGCTGAAGTTTGAAGAGCTTCAAGT 600
QY 601 CGAGTGTCAATGAGAGAGACAGCCGCGAGAGTGAACCTTCAACCCGTATGACTTGA 660
Db 601 CGAGTGTCAATGAGAGAGACAGCCGCGAGAGTGAACCTTCAACCCGTATGACTTGA 660
QY 661 CAACAAGGCGAAAGTCAACCGAGAGACATCAACAGCTTGTGACACACTTATAGAGT 720
Db 661 CAACAAGGCGAAAGTCAACCGAGAGACATCAACAGCTTGTGACACACTTATAGAGT 720
QY 721 GGTGAGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATGTGCGGGTAAAGCTCAC 780
Db 721 GGTGAGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATGTGCGGGTAAAGCTCAC 780
QY 781 CGTGGCCCGCGATGGCAGCCAGAGCAAGAGAGGCTCTTGTCAATCAAGCTGACCTGCA 840
Db 781 CGTGGCCCGCGATGGCAGCCAGAGCAAGAGAGGCTCTTGTCAATCAAGCTGACCTGCA 840
QY 841 GAGGCGCAAGGCGCCGAGAGAGACCAAGCCCACTGAGAGACTTGGGAGCTGGAGAGAA 900
Db 841 GAGGCGCAAGGCGCCGAGAGAGACCAAGCCCACTGAGAGACTTGGGAGCTGGAGAGAA 900
QY 901 GCAGGAGCCCGCTCAGGTTCCAGGGTGAACAGCCGCTGAGAGCATGTGCTGTACCA 960
```

```
Db 901 GCAGGAGCCCGCTCAGGTTCCAGGGTGAACAGCCGCTGAGAGCATGTGCTGTACCA 960
QY 961 CCATTGGTATGAGAGACATCGAGAGAGAAACCACTACTTATGATTCGCCGGAGATGA 1020
Db 961 CCATTGGTATGAGAGACATCGAGAGAGAAACCACTACTTATGATTCGCCGGAGATGA 1020
QY 1021 AAATCAAGCTCCCAATTTTGGGCTGGCTCCCTCCGTGGGCCCAAGATCAGAACTGCC 1080
Db 1021 AAATCAAGCTCCCAATTTTGGGCTGGCTCCCTCCGTGGGCCCAAGATCAGAACTGCC 1080
QY 1081 CCCCGGACCTCCAAATCCCACTGATTCGCTCCCATGAGCCGGAAGCCATCCATGCC 1140
Db 1081 CCCCGGACCTCCAAATCCCACTGATTCGCTCCCATGAGCCGGAAGCCATCCATGCC 1140
QY 1141 ACAACGAAAGCCCAAGGCGTGAACCGGCTCTCTTCACTTCTTGAACCCCAATGCC 1200
Db 1141 ACAACGAAAGCCCAAGGCGTGAACCGGCTCTCTTCACTTCTTGAACCCCAATGCC 1200
QY 1201 CAAGTCTCAGACTCCAGCAAGGCTCGGGGCAACCAAGAGGGAGCACTTGT 1260
Db 1201 CAAGTCTCAGACTCCAGCAAGGCTCGGGGCAACCAAGAGGGAGCACTTGT 1260
QY 1261 GAGGTCCCGCAAGGCCCAAGAGTGTGGTGTGGCCAGCTGCGCAGAGGGGCAAG 1320
Db 1261 GAGGTCCCGCAAGGCCCAAGAGTGTGGTGTGGCCAGCTGCGCAGAGGGGCAAG 1320
QY 1321 AAACAAGCCCTCTGTGGAGCCCGCATCTCGGTGTCTCCCTCGCGCCACTGTGCTGC 1380
Db 1321 AAACAAGCCCTCTGTGGAGCCCGCATCTCGGTGTCTCCCTCGCGCCACTGTGCTGC 1380
QY 1381 CAGCCGGGCTCTCTCCCTCCCTAGCGCCGCTCGGGCAAGAAGCAAGCAACGAGC 1440
Db 1381 CAGCCGGGCTCTCTCCCTCCCTAGCGCCGCTCGGGCAAGAAGCAAGCAACGAGC 1440
QY 1441 CAAAGAGAGCAGAGGCTGCGGGGCTGCAAGGCAACACTGAGCTCAGAGTGGCCCTGT 1500
Db 1441 CAAAGAGAGCAGAGGCTGCGGGGCTGCAAGGCAACACTGAGCTCAGAGTGGCCCTGT 1500
QY 1501 CTTGGGGGGAGACCTGCGGAGACTGCGCGCTTGTGTATGAGAGCCAGGCGCG 1560
Db 1501 CTTGGGGGGAGACCTGCGGAGACTGCGCGCTTGTGTATGAGAGCCAGGCGCG 1560
QY 1561 GCGAGCGGCTCAGAGACATGAGACCAACCAACCATGAACATCACCACTTACCA 1620
Db 1561 GCGAGCGGCTCAGAGACATGAGACCAACCAACCATGAACATCACCACTTACCA 1620
QY 1621 CTTCTACAGACATGAGACCCCTCCCAAGGCGCCCACTGCAATGAAAGACCCCA 1680
Db 1621 CTTCTACAGACATGAGACCCCTCCCAAGGCGCCCACTGCAATGAAAGACCCCA 1680
QY 1681 CCGGACACCAAGGACATTTATTTCTATTTATTTATTTATTTATTTATTTATTTAT 1740
Db 1681 CCGGACACCAAGGACATTTATTTCTATTTATTTATTTATTTATTTATTTATTTAT 1740
QY 1741 TAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1800
Db 1741 TAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1800
QY 1801 CACAGAACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1859
Db 1801 CACAGAACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1859
```

RESULT 2
ID AAA63925 standard; DNA; 1438 BP.

AC AAA63925;

XX 04-DEC-2000 (first entry)

XX DNA encoding protein related to Drosophila naked cuticle polypeptide.

XX Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
KW gene therapy; Nkd defect; cancer; 88.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 8..1420
FT /tag= a
FT /product= "naked cuticle polypeptide"
XX
PN MO200049034-A1.
XX
PD 24-AUG-2000.
XX
PF 17-FEB-2000; 2000MO-US004188.
XX
PR 17-FEB-1999; 99US-0120646P.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Scott M, Zeng W, Wharton K;
XX
DR MPI: 2000-571967/53.
XX P-PSDB; AAB08216.
XX
XX An isolated nucleic acid molecule useful for analyzing (genetic
PT predisposition to) a disease state and for therapeutic purposes e.g.
PT treatment of cancer comprises a sequence encoding a naked cuticle
PT protein.
XX
XX Claim 3; Page 46-48; 58pp; English.
XX
XX The present sequence encodes a protein related to the Drosophila Nkd
CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide
CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the
CC regulation of Wnt signal potency, duration or distribution. The Nkd
CC polynucleotides can be used for identifying homologous or related
CC proteins, to modulate the expression or function of Nkd polynucleotides, and
CC in studying associated physiological pathways. Nkd polynucleotides can
CC also be used in gene therapy to treat disorders associated with Nkd
CC defects. They may also be used for therapeutic purposes e.g. treatment of
CC cancer.
XX
XX Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;
SQ
Query Match 71.9%; Score 1336; DB 3; Length 1438;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 217 CCCACAGTGGGAACTTCACTCAAGCCGCGCGCTGTGCAAGCGAGGAGAGCCC 276
DB 1 CCCACAGTGGGAACTTCACTCAAGCCGCGCGCTGTGCAAGCGAGGAGAGCCC 60
QY 277 GGAAGTGACAGTTGGCCCTGAGCCGCTGCTGGGCTCGAAGGGCATCGAGAGTGGAT 336
DB 61 GGAAGTGACAGTTGGCCCTGAGCCGCTGCTGGGCTCGAAGGGCATCGAGAGTGGAT 120
QY 337 CGGAGACAGCGCTGCGCGCGCGGTGTCTGGGACCCCGACAGCTGGCGTGGCGGAC 396
DB 121 CGGAGACAGCGCTGCGCGCGGTGTCTGGGACCCCGACAGCTGGCGTGGCGGAC 180
QY 397 CATAGCCGGAAGACCCGAGAGCTCTGGGCGAGCTGTGTAAGAGACAGCTCAGCGAGGA 456
DB 181 CATAGCCGGAAGACCCGAGAGCTCTGGGCGAGCTGTGTAAGAGACAGCTCAGCGAGGA 240
QY 457 AAGAGAGAGAGCATTTTGGCTGGAAGTGGCTTGGCTCTTGAAGAGACTGAAGGCTGGG 516
DB 241 AAGAGAGAGAGCATTTTGGCTGGAAGTGGCTTGGCTCTTGAAGAGACTGAAGGCTGGG 300
QY 517 CACGGAGATGAGAGAGAGATGAGAGAGTGAAGGAACTTGGCCAGGCTCCAAGAGCA 576
|||||

DB 301 CAGCGAGATGAGAGAGATGAGAGAGTGAAGCAACCTGGCCAGGCTCCAAGAGCA 360
QY 577 GCTGAAGTTGAAGAGCTCCAGTGCGAGCTGTCCATGAGAGAGACCGCGCAGAGTG 636
DB 361 GCTGAAGTTGAAGAGCTCCAGTGCGAGCTGTCCATGAGAGAGAGACCGCGCAGAGTG 420
QY 637 GACCTTCAACCTGTATGACTTTTGAACAACAACGGCAAGGTCAACCGAGAGACATCACAG 696
DB 421 GACCTTCAACCTGTATGACTTTTGAACAACAACGGCAAGGTCAACCGAGAGACATCACAG 480
QY 697 CTGTGTCACACCATCTATGAGGTGTGTGACTCTCTGTCAACCACTCCCCAATCCAG 756
DB 481 CTGTGTCACACCATCTATGAGGTGTGTGACTCTCTGTCAACCACTCCCCAATCCAG 540
QY 757 CAAGATCTCGGGGTAAAGCTCAACCGTGCCCCGATGCGACCGCAAGAGAGAGCGT 816
DB 541 CAAGATCTCGGGGTAAAGCTCAACCGTGCCCCGATGCGACCGCAAGAGAGAGCGT 600
QY 817 CTTTGTCAATCAGAGCTGACCTGTGAGAGCGGCAAGCCCCGAGAGACCAAGCCACTGA 876
DB 601 CTTTGTCAATCAGAGCTGACCTGTGAGAGCGGCAAGCCCCGAGAGACCAAGCCACTGA 660
QY 877 GGAAGCTGCGGAGCTGGGAGAGAGAGAGAGAGCGAGCCGCTCAGGTTCCAGGGTGAACGCG 936
DB 661 GGAAGCTGCGGAGCTGGGAGAGAGAGAGAGAGAGCGAGCCGCTCAGGTTCCAGGGTGAACGCG 720
QY 937 CTTGAGACAGTCTGGCTCTACCAACCATTTGATGAGAGACATCGAGAGAGAAACA 996
DB 721 CTTGAGACAGTCTGGCTCTACCAACCATTTGATGAGAGACATCGAGAGAGAAACA 780
QY 997 CTACTTAAGATCTGCGCGGATAGAAAACCTACAGTCCCAATTTTGGGCTGCTCCCTTC 1056
DB 781 CTACTTAAGATCTGCGCGGATAGAAAACCTACAGTCCCAATTTTGGGCTGCTCCCTTC 840
QY 1057 CGTGCCCAAGATCAGAACTGCCCCCGCACCTCCCAATCCCACTGTGCTGCCCA 1116
DB 841 CGTGCCCAAGATCAGAACTGCCCCCGCACCTCCCAATCCCACTGTGCTGCCCA 900
QY 1117 TGAAGCGGAAGCATTCACATCCACACCGAAAGCCCAAGCGGTGAGCCCGCTCTCTT 1176
DB 901 TGAAGCGGAAGCATTCACATCCACACCGAAAGCCCAAGCGGTGAGCCCGCTCTCTT 960
QY 1177 CCACTTCTTGAACACCCCAATGCGCAAGTCTGAGAGTCTCCACCAACGGTCCGGGAC 1236
DB 961 CCACTTCTTGAACACCCCAATGCGCAAGTCTGAGAGTCTCCACCAACGGTCCGGGAC 1020
QY 1237 CCAAGACGGGAGCAAGCATTTGTGAGGTCCCCCAAGGCCAGGAGAGAGTGGGTGT 1296
DB 1021 TCAAGACGGGAGCAAGCATTTGTGAGGTCCCCCAAGGCCAGGAGAGAGTGGGTGT 1080
QY 1297 GGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGGACCCGCAATCCCTGGGT 1356
DB 1081 GGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGGACCCGCAATCCCTGGGT 1140
QY 1357 GTCCCCCTCGGCCCACTGTGCTGTCAGAGCCCGGCTCTCTCCCTTACCCCCCTCGG 1416
DB 1141 GTCCCCCTCGGCCCACTGTGCTGTCAGAGCCCGGCTCTCTCCCTTACCCCCCTCGG 1200
QY 1417 GCACAAGAGCAACAAGCACCGAGCCCAAGAGAGACCAAGAGGCTGCGGAGGCTGCAAGG 1476
DB 1201 GCACAAGAGCAACAAGCACCGAGCCCAAGAGAGACCAAGAGGCTGCGGAGGCTGCAAGG 1260
QY 1477 ACCACTGGCTCAAGTGGCCCTGTCTGTGGGCGGAGACCTGCGAGAGCTGCCGCTT 1536
DB 1261 ACCACTGGCTCAAGTGGCCCTGTCTGTGGGCGGAGACCTGCGAGAGCTGCCGCTT 1320
QY 1537 GGTGTGTATGAGAGCAAGGCGGGGAGACCGGCTCCAGAGCAAGAGACCAACCAACCA 1596
DB 1321 GGTGTGTATGAGAGCAAGGCGGGGAGACCGGCTCCAGAGCAAGAGACCAACCAACCA 1380
QY 1597 TGAACATCAACCACTTACCACTTCAACAGACATAGAGCCCTCCCAAGGCGCC 1654
DB 1381 TGAACATCAACCACTTACCACTTCAACAGACATAGAGCCCTCCCAAGGCGCC 1438
|||||

RESULT 3
AAL03214
ID AAL03214 standard; DNA; 2379 BP.
XX
XX AAL03214;
AC
XX 21-NOV-2001 (first entry)
DT
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 5902.
DE
XX Human reproductive system related antigen; reproductive system disorder;
KM cancer; gene therapy; ds.
XX
XX Homo sapiens.
OS
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232402P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234397P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244674P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249254P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249259P.
PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 XX Isolated nucleic acid molecule encoding a reproductive system antigen is
 XX used in preventing, treating or ameliorating a medical condition.
 PT
 PS Disclosure; SEQ ID NO 5902; 1297bp + Sequence Listing; English.
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention
 XX
 SQ Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;
 Query Match 42.1%; Score 783; DB 4; Length 2379;
 Best Local Similarity 100.0%; Pred. No. 2.5e-284;
 Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1046 GGGTCCCTCCCTGCGCCGAGGAGTCAAGTCACTGCCCCCGGACCTTCAATCCCACTCGA 1105
 DB 840 GGGTCCCTCCCTGCGCCGAGGAGTCAAGTCACTGCCCCCGGACCTTCAATCCCACTCGA 899
 QY 1106 TCTCGTCCCATGAGCGGAGGAGTCAATCCCACTGCCCCCGGAGGCGGTGGAC 1165
 DB 900 TCTCGTCCCATGAGCGGAGGAGTCAATCCCACTGCCCCCGGAGGCGGTGGAC 959
 QY 1166 CCGGCTCTCTTCACTTCTTGAACCCCAATGCGCAAGGTCTCAGAGTTCAGCAACGG 1225
 DB 960 CCGGCTCTCTTCACTTCTTGAACCCCAATGCGCAAGGTCTCAGAGTTCAGCAACGG 1019
 QY 1226 CTCGGGGGACCCAGAGCGGAGGAGCACTTTGTGAGTTCCTCCCAAGGCCCAAGGCGAAG 1285
 DB 1020 CTCGGGGGACCCAGAGCGGAGGAGCACTTTGTGAGTTCCTCCCAAGGCCCAAGGCGAAG 1079
 QY 1286 AGTGTGGTGTGGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1345
 DB 1080 AGTGTGGTGTGGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1139
 QY 1346 ATCCCTGCGGTGTCTCCCTCTCCGCACTTGTGAGTTCAGAGGCGGCTCTCTCCCTCCCTA 1405
 DB 1140 ATCCCTGCGGTGTCTCCCTCTCCGCACTTGTGAGTTCAGAGGCGGCTCTCTCCCTCCCTA 1199
 QY 1406 GCGGCTCTGCGGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAG 1465
 DB 1200 GCGGCTCTGCGGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAG 1259
 QY 1466 GCGCTGAGGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAG 1525
 DB 1260 GCGCTGAGGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAG 1319
 QY 1526 CTGCGGCTCTGCGGTGTGTAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAG 1585
 DB 1320 CTGCGGCTCTGCGGTGTGTAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAG 1379

QY 1586 CACCAACGATGAACATCAACCATTAACACACTTCTACAGAGATGAGCCCTCC 1645
 DB 1380 CACCAACGATGAACATCAACCATTAACACACTTCTACAGAGATGAGCCCTCC 1439
 QY 1646 CAGAGGCCCCACCTGCGCATATGAAGAGCCCAACCCCGAGACAGCAAGCATTAATT 1705
 DB 1440 CAGAGGCCCCACCTGCGCATATGAAGAGCCCAACCCCGAGACAGCAAGCATTAATT 1499
 QY 1706 CTATTAATTAATGTTATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1765
 DB 1500 CTATTAATTAATGTTATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1559
 QY 1766 ATTAGTACCTTACATGATGAAGATTAATGAAGAGCAAGAGCAAGAGCAAGAGCAAGAG 1825
 DB 1560 ATTAGTACCTTACATGATGAAGATTAATGAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAG 1619
 QY 1826 GTT 1828
 DB 1620 GTT 1622
 RESULT 4
 AAL03216
 ID AAL03216 standard; DNA; 2379 BP.
 XX
 AC AAL03216;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen DNA SEQ ID NO: 5904.
 XX
 KM Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ds.
 OS Homo sapiens.
 XX
 PN NO200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001MO-US001339.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 5904; 1297bp + Sequence Listing; English.
PS
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;
Query Match 42.1%; Score 783; DB 4; Length 2379;
Best Local Similarity 100.0%; Pred. No. 2,5e-284;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGCCTCCCTCCGTCGCGCCGGAAGTCAGACTGCCCCCGGACCTTCATCCACCGA 1105
DB 840 GGCCTCCCTCCGTCGCGCCGGAAGTCAGACTGCCCCCGGACCTTCATCCACCGA 899
QY 1106 TCTCGCTCCATGAGCCGGAAGCATCCATCCACCGAAGCCCAAGGGGTGAC 1165
DB 900 TCTCGCTCCATGAGCCGGAAGCATCCATCCACCGAAGCCCAAGGGGTGAC 959
QY 1166 CCGGCTCTCTTCACTTCTTGAACCCCAATGCGCAAGGTTTCAGAGCTTCAGCAACGG 1225
DB 960 CCGGCTCTCTTCACTTCTTGAACCCCAATGCGCAAGGTTTCAGAGCTTCAGCAACGG 1019
QY 1226 CTCGGGGGACCCAGGACGGGAGCAAGCACTTTGTGAGTCCCCCAAGGGCCAGGGCAAG 1285
DB 1020 CTCGGGGGACCCAGGACGGGAGCAAGCACTTTGTGAGTCCCCCAAGGGCCAGGGCAAG 1079
QY 1286 AGTGTGGGTGTGGGACAGTGGCCAGAGGGGCAAGAAAGCCCTCTGGGACCCGCC 1345

Dd		1080	AGTGTGGGTGTGAGGCACATGTGGCCAGAGGGGGAATAAACAGCCCCCTCTGGAAACCGGCC	1139
Oy		1346	ATCCTCGGTTGTTCCTCCCTCGGCCCACTGTGTCAGACCCTGGCTCTCCCTCCCTTA	1405
Dd		1140	ATCCCTGGGTTGTCCCCTCCGCGCCACTGGGTGCACCCGGGCTCTCTCCCTCCCTA	1199
Oy		1406	GCCCCCTTCGGGGCANMAAGAACAAGCACCGAGGCAAGAGAGAGCAGACAGAGGCTGCGCG	1465
Dd		1200	GCCCCCTTCGGGGCANMAAGAACAAGCACCGAGGCAAGAGAGCAGACAGAGGCTGCGCG	1259
Oy		1466	GAGCTGACGACACACTGAGCTCAGGTGAGCTGTCTGTGGGCGGAGACACTGCGGGAG	1525
Dd		1260	GAGCTGACGACACACTGAGCTCAGGTGAGCTGTCTGTGGGCGGAGACACTGCGGGAG	1319
Oy		1526	CTGCCCCGCTTGTTGTGTGTATGAGAGCCAGGCGGGCAGCCGGTCCAGAGACATGAGCAC	1585
Dd		1320	CTGCCCCGCTTGTTGTGTGTATGAGAGCCAGGCGGGCAGCCGGTCCAGAGACATGAGCAC	1379
Oy		1586	CACCAACCAACATGAACATCACCACTTAACCAACCATCTTACACAGACATAGACCCCTCC	1645
Dd		1380	CACCAACCAACATGAACATCACCACTTAACCAACCATCTTACACAGACATAGACCCCTCC	1439
Oy		1646	CCAGGGCCCCACCCCTGGCATATGAGAGACCCCAACCCCGGACACCAAGAGGATTATTAT	1705
Dd		1440	CCAGGGCCCCACCCCTGGCATATGAGAGACCCCAACCCCGGACACCAAGAGGATTATTAT	1499
Oy		1706	CTATTAAATTATTTGTTATTATGATGATTAATTTGTTATTAAATTATTTGTTACTGCACATA	1765
Dd		1500	CTATTAAATTATTTGTTATTATGATGATTAATTTGTTATTAAATTATTTGTTACTGCACATA	1559
Oy		1766	ATTTAGCTAGCCCTACATGTAGAGATCTATGAGAAACACAGAACTTAACTTTATTATAT	1825
Dd		1560	ATTTAGCTAGCCCTACATGTAGAGATCTATGAGAAACACAGAACTTAACTTTATTATAT	1619
Oy		1826	GTT 1828	
Dd		1620	GTT 1622	
RESULT 5				
AAL51525				
ID AAL51525	standard; DNA; 590 bp.			
XX AC				
XX AAL51525;				
XX DT	24-APR-2003	(first entry)		
XX Dd				
XX Xx				
Xx Human Nkd (hnkd) gene exon 11 (coding region).				
Kw Human; dg; gene therapy; Nkd; hnkd; Wnt signalling pathway; cancer;				
Km colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;				
Rn promoter.				
OS Homo sapiens.				
XX MO200292832-A2.				
XX PD	21-NOV-2002.			
XX PF	27-NOV-2001; 2001WO-US044092.			
XX PR	27-NOV-2000; 2000US-0252884P.			
PR 16-MAY-2001; 2001US-0291109P.				
PR 01-OCT-2001; 2001US-0325571P.				
XX PA	(CHIR) CHIRON CORP.			
XX PI	Rohan M, Chan V, Yan D;			
XX DR	WPI; 2003-129303/12.			
XX LT	New human and non-human primate homologues of Nkd protein, and Nkd genes,			

PT		useful for treating cancer involving aberrant Wnt signaling, e.g. colon cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX		
PS		Example 3; Fig 7; 99p; English.
XX		
CC		The invention comprises the amino acid and coding sequence of the human Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC		and protein sequences are useful for the treatment of cancer involving
CC		aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC		cancer and breast cancer). The present DNA sequence represents exon 11
CC		(coding region) from the human Nkd gene
XX		
SQ		Sequence 590 BP; 119 A; 234 C; 159 G; 78 T; 0 U; 0 Other;
QY		
Db		
Query Match	31.7%; Score 590; DB 10; Length 590;	
Best Local Similarity	100.0%; Pred. No. 9.1e-212;	
Matches 590; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1047	GCTCCCCCTTCGTCGGCCAGAAATCAGAATCGCCCCCGGCACCTCCCAATGCCACTGCAT 1106
Db	1	GCTCCCCCTTCGTCGGCCAGAAATCAGAATCGCCCCCGGCACCTCCCAATGCCACTGCAT 60
QY	1107	CTCGCTCCCATGAGCGCGGAAGCCATTCACATCCCAACCGAAGCCCAAAGCGGTGAGC 1166
Db	61	CTCGCTCCCATGAGCGCGAGCCATTCACATCCCAACCGAAGCCCAAAGCGGTGAGC 120
QY	1167	CGGCTCTCTTCCACTTCTTTGACAACCCCAATCGCCAAAGTCTCAAGCTCCAGCAACGGC 1226
Db	121	CGGCTCTCTTCCACTTCTTTGACAACCCCAATCGCCAAAGTCTCAAGAGTCTCAAGCAACGGC 180
QY	1227	TCCGGGAGCACCCAGAGACGGAGCAAGCATTTTGAGAGTCCCCCAAGGCCCAAGGCAAGA 1286
Db	181	TCCGGGAGCACCCAGAGACGGAGCAAGCATTTTGAGAGTCCCCCAAGGCCCAAGGCAAGA 240
QY	1287	GTCGTGGGTGTGGGCGACGTGGCCAGAGGGGCGCAAAAACAAGCCCCTTGGGACCCGGCA 1346
Db	241	GTCGTGGGTGTGGGCGACGTGGCCAGAGGGGCGCAAAAACAAGCCCCTTGGGACCCGGCA 300
QY	1347	TCCCTGTGGGTGTCCCCCTCGCCCACTGTGCTGCAGCCCGACCCTCTCCCTCCCTTAG 1406
Db	301	TCCCTGTGGGTGTCCCCCTCGCCCACTGTGCTGCAGCCCGACCCTCTCCCTCCCTTAG 360
QY	1407	CCCCCTTCGGGACAAAGAACCAAGCACCGAGCCCAAGAGAGCCACAGGGCTGCGGG 1466
Db	361	CCCCCTTCGGGACAAAGAACCAAGCACCGAGCCCAAGAGAGCCACAGGGCTGCGGG 420
QY	1467	GCCGTGAGGACCACTGAGCTCAGTGGCCCTGTCTCTGGGGCGGGACCACTTGGGGAGC 1526
Db	421	GCCGTGAGGACCACTGAGCTCAGTGGCCCTGTCTCTGGGGCGGGACCACTTGGGGAGC 480
QY	1527	TGCCCGCTTGTGTGTATGAGAGCGAGCGGGGACCGGTCTCAAGACATGAGCAAC 1586
Db	481	TGCCCGCTTGTGTGTATGAGAGCGAGCGGGGACCGGTCTCAAGACATGAGCAAC 540
QY	1587	ACCAACCAACATGAACATCACCAACCATTAACCACTTCAACAGACATTAG 1636
Db	541	ACCAACCAACATGAACATCACCAACCATTAACCACTTCAACAGACATTAG 590
RESULT 6		
AAS92082		
ID	AAS92082 standard; cDNA; 1561 BP.	
AC	AAS92082;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	DNA encoding novel human diagnostic protein #27886.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KM	food supplement; medical imaging; diagnostic; genetic disorder; sv.	
XX		
OS	Homo sapiens.	

[illegible]

Db	301	GGGAGAAAGACGAGGAGCCCCGCTCAGTTCCAGGTGACAGCCGCTGGAGCAGTCTG	360
Qy	951	GCTGCTACACCACTTTCGCTAGATGAGAACATTCGAGAGGAAACCACTTACTTCTCG	1011
Db	361	GCTGCTACCAACCATTCGCTAGATGAGAACATTCGAGAGGAAACCACTTACTTCTCG	420
Qy	1011	CCGGGATGAAAACTCACGCTCCCAATTTGGGCGCTG	1046
Db	421	CCGGGATGAAAACTCACGCTCCCAATTTGGGCGCTG	456
RESULT 7			
ID	ACH21474	standard; cDNA, 427 BP.	
XX	ACH21474;		
XX	13-OCT-2003	(first entry)	
XX	DT		
DE	Human adult liver cDNA #1086.		
XX	Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;		
KW	genome mapping; biodiversity; genetic disorder.		
XX			
OS	Homo sapiens.		
XX	US2003073623-A1.		
XX	17-APR-2003.		
PF	30-JUL-2001; 2001US-00918995.		
XX	30-JUL-2001; 2001US-00918995.		
XX	30-JUL-2001; 2001US-00918995.		
PA	(DRWA/) DRMANAC R T.		
PA	(LAEA/) LABAT I.		
PA	(STAC/) STRACHE-CRAIN B.		
PA	(DICK/) DICKSON M C.		
XX	(JONE/) JONES L W.		
PI	Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;		
XX	WPI, 2003-615964/58.		
XX	New polynucleotide sequences obtained from various cDNA libraries, useful		
PT	as hybridization probes, as oligomers for PCR, for chromosome and gene		
PT	mapping, in the recombinant production of protein, or in generating		
PT	antisense DNA or RNA.		
PS	Claim 1; SEQ ID NO 8686; 44pp; English.		
XX	The invention relates to an isolated polynucleotide comprising any one of		
CC	38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was		
CC	determined by the technique of SBH (sequencing by hybridisation). Also		
CC	included is a purified polypeptide comprising a sequence corresponding to		
CC	a reading frame of the novel polynucleotide. The nucleic acid sequences		
CC	are useful in diagnostics as expressed sequence tags (EST) for		
CC	identifying expressed genes or for physical mapping of the human genome,		
CC	in forensics, in assessing biodiversity, or in identifying mutations		
CC	responsible for genetic disorders and other traits. The nucleotide		
CC	sequences are also useful as hybridisation probes, as oligomers for PCR,		
CC	for chromosome and gene mapping, in the recombinant production of		
CC	protein, or in generating antisense DNA or RNA. The purified polypeptide		
CC	is useful for generating antibodies specific for it. The present sequence		
CC	is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data		
CC	for this patent did not form part of the printed specification, but was		
CC	obtained in electronic format directly from USPTO at		
XX	segdata.uspto.gov/sequence.html?DocID=20030073623		
XX			
XX	Sequence 427 BP; 84 A; 177 C; 112 G; 53 T; 0 U; 1 Other;		

Query Match 20.3%; Score 377; DB 9; Length 427;

	Best Local Similarity	100.0%;	Pred. No. 8.9e-132;	Mismatches	0;	Indels	0;	Gaps	0;	
	Matches	377;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Oy	1076	CTGCCCCCCCAGCCTTCCATTCACACTGCATCTCGCTCCCATGTAGCCGGAAAGCCATCCAC	1135							
Dd	4	CTGCCCCCCCGCACCTCCMATCCCACTGATCTTGCTCTCCATTGACCAGGAAAGCCATCCAC	63							
Oy	1136	ATCCCAACCGAAGAGCCCAAGGCGGTGAGCCCGGCTCTTCCAATTCTTTGACACCCCA	1195							
Dd	64	ATCCCAACCGAAGAGCCCAAGGCGGTGAGCCCGGCTCTTCCAATTCTTTGACACCCCA	123							
Oy	1196	ATGCCCAAGTCTCTAGAGCTCCAGCAACGCTCTCGGGGCAACCCAGAGCGGAAGCAAC	1255							
Dd	124	ATCCCAAGTCTCTAGAGCTCCAGCAACGCTCTCGGGGCAACCCAGAGCGGAAGCAAC	183							
Oy	1256	TTTGTGAGTCTCCCAAGGCGCCAGGAGCAAGTGTGGGTGTGGGCAACGTGGCAAGAGG	1315							
Dd	184	TTTGTGAGTCTCCCAAGGCGCCAGGAGCAAGTGTGGGTGTGGGCAACGTGGCAAGAGG	243							
Oy	1316	GCAAGAAAACAAGCCCTCTGTGGAACCCGACATCCCTGTGGTGTCCCTCCGCCACCTG	1375							
Dd	244	GCAAGAAAACAAGCCCTCTGTGGAACCCGACATCCCTGTGGTGTCCCTCCGCCACCTG	303							
Oy	1376	GCTGCCAGCCCGGCGCTCTCTCTCCCTTACGCCCGCTCGGGCAACAGAACAAGCAC	1435							
Dd	304	GCTGCCAGCCCGGCGCTCTCTCTCCCTTACGCCCGCTCGGGCAACAGAACAAGCAC	363							
Oy	1436	CGAGCCAAAGAGAGCCA 1452								
Dd	364	CGAGCCAAAGAGAGAGCCA 380								
RESULT 8										
AAS78752										
ID	AAS78752	standard; cDNA; 1743 BP.								
XX	AAS78752;									
AC	13-FEB-2002	(first entry)								
DT										
XX										
DE	DNA encoding novel human diagnostic protein #14556.									
XX										
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;									
KM	Food supplement; medical imaging; diagnostic; genetic disorder; ss.									
OS	Homo sapiens.									
XX										
PN	WO200175067-A2.									
XX										
PD	11-OCT-2001.									
XX										
PF	30-MAR-2001; 2001WO-US008631.									
XX										
PR	31-MAR-2000; 2000US-00540217.									
XX										
PR	23-AUG-2000; 2000US-00649167.									
XX										
PA	(HYSE-) HYSEQ INC.									
XX										
PI	Dermanac RT, Liu C, Tang YT;									
XX										
DR	WPI; 2001-639362/73.									
XX										
DR	P-PSDB; ABG14565.									
XX										
PT	New isolated polynucleotide and encoded polypeptides, useful in									
XX	diagnostics, forensics, gene mapping, identification of mutations									
PT	responsible for genetic disorders or other traits and to assess									
XX	biodiversity.									
XX										
PS	Claim 1; SEQ ID NO 14556; 103bp; English.									
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)									
CC	sequences. (I) is useful as hybridisation probes, polymerase chain									
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,									

Query	Match	Similarity	Score	DB	Length
Best Local	Similarity 100.0%; Pred. No. 3.5e-72;				1743;
Matches	218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1047 GCTCCCTTCCTCGTGGCCAGAGTCAAGACTGCCCCCGCAGCCTCCATCCCACTGGAT				1106
DB	332 GCTCCCTTCCTCGTGGCCAGAGTCAAGACTGCCCCCGCAGCCTCCATCCCACTGGAT				391
QY	1107 CTCGCTCCCATGAGCCGGAAAGCCATCCATCCCAACCGAAAGCCCAAGGCGCTGAGAC				1166
DB	392 CTCGCTCCCATGAGCCGGAAAGCCATCCATCCCAACCGAAAGCCCAAGGCGCTGAGAC				451
QY	1167 CGGCGCTCCTTCCATCTTCTTGACACCCCATTCGCAAGGTCTCAGAGCTCCAGCAAGGC				1226
DB	452 CGGCGCTCCTTCCATCTTCTTGACACCCCATTCGCAAGGTCTCAGAGCTCCAGCAAGGC				511
QY	1227 TCCGGGGCACCAGGAGCGGAGCAAGCATTGTGAGG				1264
DB	512 TCCGGGGCACCAGGAGCGGAGCAAGCATTGTGAGG				549
RESULT 9					
AAS92081					
ID	AAS92081	standard; cDNA; 1743 BP.			
XX	AC	AAS92081;			
XX	DT	13-FEB-2002 (first entry)			
XX	DE	DNA encoding novel human diagnostic protein #27885.			
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
XX	OS	food supplement; medical imaging; diagnostic; genetic disorder; ss.			
XX	PN	MO200175067-A2.			
XX	PD	11-OCT-2001.			
XX	PF	30-MAR-2001; 2001MO-US008631.			
XX	PR	31-MAR-2000; 2000US-00540217.			
XX	PR	23-AUG-2000; 2000US-00649167.			
XX	PA	(HYSE-) HYSEQ INC.			
XX	PI	Drmanac RT, Liu C, Tang YT;			
XX	DR	WPI; 2001-639362/73.			
XX	DR	P-PSDB; ABG27894.			

XX Human genome derived single exon probe #7990.
 DE Human; probe; ss; gene expression; single exon probe; microarray;
 XX alternative splicing event; genomic alteration.
 KW Homo sapiens.
 XX US2003194704-A1.
 PN 16-OCT-2003.
 PD 03-APR-2002; 2002US-00029386.
 PF 03-APR-2002; 2002US-00029386.
 PR (PENN/) PENN S G.
 PA (PENN/) RANK D R.
 PA (HANK/) HANZEL D K.
 XX Penn SG, Rank DR, Hanzel DK;
 PI WPI; 2004-119264/12.
 DR New human genome-derived single exon nucleic acid probes useful for human
 XX gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 PT Claim 15; SEQ ID NO 7990; 80pp; English.
 XX The invention relates to a nucleic acid probe for measuring human gene
 XX expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20030194704
 XX Sequence 598 BP; 84 A; 232 C; 162 G; 120 T; 0 U; 0 Other;
 SQ
 Query Match 7.4%; Score 137; DB 12; Length 598;
 Best Local Similarity 100.0%; Pred. No. 1.1e-41;
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 AGGTGACAGCTTCCGCGTGTGAGCCGCTGCTGGAGGAGCATCGAGAGTGCATCGG 339
 |||||
 DB 228 AGGTGACAGCTTCCGCGTGTGAGCCGCTGCTGGAGGAGCATCGAGAGTGCATCGG 287
 QY 340 GAGACAGCGCTGCCCGGCGGTGTCTTCGGGAGCCCGACAGCTCGGTGGCGGACCAT 399
 |||||
 DB 288 GAGACAGCGCTGCCCGGCGGTGTCTTCGGGAGCCCGACAGCTCGGTGGCGGACCAT 347
 QY 400 AGGCCGAGAGCACCCTGGG 416
 |||||
 DB 348 AGGCCGAGAGCACCCTGGG 364
 RESULT 14
 ID ABQ73820 standard; cDNA; 1950 BP.
 XX ABQ73820;
 AC 07-OCT-2002 (first entry)
 DT Human colon specific nucleic acid (CSNA) SEQ ID NO:126.
 DE Human; colon specific nucleic acid; colon specific polypeptide; CSP;
 KW CSNA; colon specific gene; CSG; colon cancer; gene therapy; vaccine;
 KW cytoskeletal; gene; ss.
 XX Homo sapiens.
 OS WO200248370-A2.
 PN 20-JUN-2002.
 PD 30-OCT-2001; 2001WO-US051341.
 PF 31-OCT-2000; 2000US-0244717P.
 PR (DIAD-) DIADEXUS INC.
 PA Sun Y, Reclipon H, Ghosh MG, Liu C;
 PI WPI; 2002-583520/62.
 DR Colon specific polypeptides and polynucleotides useful for detecting,
 XX diagnosing, monitoring, treating, staging and predicting cancers in
 PT humans having cancer and non-cancerous colon disease.
 PT Claim 1; Page 199; 243pp; English.
 PS ABQ73695 to ABQ73841 represent human colon specific nucleic acid (CSNA)
 CC sequences, and ABP51826 to ABP51928 represent human colon specific
 CC polypeptide (CSP) sequences from the present invention. CSNA and CSP
 CC sequences have cytostatic activity, and can be used in gene therapy, for
 CC antitumor therapy and in vaccines. CSNA and CSP sequences can be used for
 CC diagnosing and monitoring the presence and metastases of colon cancer in
 CC a patient, by determining an amount of CSP or CSNA in a sample of a
 CC patient, and comparing it to the amount of colon specific marker in a
 CC normal control, where a difference in the amount of the nucleic acid or
 CC the polypeptide in the sample compared to that of normal control is
 CC associated with presence of colon cancer. CSP and CSNA sequences can be
 CC used for producing engineered colon tissue for treatment and research.
 CC CSNA sequences are useful for producing transgenic animals and cells and
 CC also in gene therapy
 XX Sequence 1950 BP; 459 A; 494 C; 530 G; 467 T; 0 U; 0 Other;
 SQ
 Query Match 7.4%; Score 137; DB 6; Length 1950;
 Best Local Similarity 100.0%; Pred. No. 8.8e-42;
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 281 GGTGACAGCTTCCCGTGTGAGCCGCTGCTGGAGGAGCATCGAGAGTGCATCGG 340
 |||||
 DB 838 GGTGACAGCTTCCCGTGTGAGCCGCTGCTGGAGGAGCATCGAGAGTGCATCGG 897

Search completed: December 29, 2004, 21:21:14
Job time: 1271 secs

QY 341 AGACAGCGCTGCGCGCGGCTGTCTCGGAGCCCGAGACGCTGCGGTTGGCGGACCATTA 400
DB 898 AGACAGCGCTGCGCGCGGCTGTCTCGGAGCCCGAGACGCTGCGGTTGGCGGACCATTA 957
QY 401 GCGCGAAGCACCAGGGA 417
DB 958 GCGCGAAGCACCAGGGA 974

RESULT 15

AL51518
ID AL51518 standard; DNA; 134 BP.

AC AAL51518;

DT 24-APR-2003 (first entry)

DE Human Nkd (hNkd) gene exon 4 (coding region).

KW Human; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;
KM colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
XX promoter.

OS Homo sapiens.

XX WO200292832-A2.

PN 21-NOV-2002.

PD 27-NOV-2001; 2001WO-US044092.

XX 27-NOV-2000; 2000US-0252884P.

PR 16-MAY-2001; 2001US-0291109P.

PR 01-OCT-2001; 2001US-0325571P.

XX (CHIR) CHIRON CORP.

PI Rohan M, Chan V, Yan D;

DR WPI; 2003-129303/12.

PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
PT useful for treating cancer involving aberrant Wnt signaling, e.g. colon
PT cancer, head and neck cancer, ovarian cancer, or breast cancer.

PS Example 3; Fig 7; 99p; English.

CC The invention comprises the amino acid and coding sequence of the human
CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence represents exon 4
CC (coding region) from the human Nkd gene

XX Sequence 134 BP; 21 A; 40 C; 55 G; 18 T; 0 U; 0 Other;

Query Match 7.2%; Score 134; DB 10; Length 134;

Best Local Similarity 100.0%; Pred. No. 1.8e-40; Mismatches 0; Gaps 0;

DB 282 GTGACAGCTTGGCGCGGCTGCTCGGAGCCCGAGACGCTGCGGTTGGCGGACCATTA 341
1 GTGACAGCTTGGCGCGGCTGCTCGGAGCCCGAGACGCTGCGGTTGGCGGACCATTA 60

QY 342 GACAGGCGTGGCGGCGGCTGCTCGGAGCCCGAGACGCTGCGGTTGGCGGACCATTA 401
DB 61 GACAGGCGTGGCGGCGGCTGCTCGGAGCCCGAGACGCTGCGGTTGGCGGACCATTA 120

QY 402 GCGGAAGCACCAGG 415
DB 121 GCGGAAGCACCAGG 134

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 20:56:19 ; Search time 212 Seconds

(Without alignments)
6232.814 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859

Sequence: 1 ggaatcgccctctatacgcg.....aaaaaaaaaagcgccgc 1859

Scoring table: ~~OEIGENUC~~
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B COMB.seq: *
3: /cgn2_6/prodata/1/ina/5A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCMB COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1336	71.9	1438	4	US-09-506-066E-5
2	53	2.9	3768	4	US-10-101-464A-949
3	50	2.7	1731	4	US-09-506-066E-3
4	45	2.4	45	4	US-09-686-583B-65
5	45	2.4	45	4	US-09-672-725C-18
6	31	1.7	47	2	US-08-778-494B-114
7	31	1.7	50	1	US-08-381-572-20
8	31	1.7	50	1	US-08-592-820-20
9	31	1.7	50	4	US-09-129-603-3
10	31	1.7	51	2	US-08-582-562A-8
11	31	1.7	51	2	US-08-778-494B-8
12	31	1.7	51	2	US-08-859-998-1373
13	31	1.7	51	3	US-09-225-928-1373
14	31	1.7	51	4	US-09-225-928-1373
15	31	1.7	52	3	US-08-618-100B-9
16	31	1.7	52	4	US-09-807-784B-8
17	31	1.7	55	2	US-08-582-562A-16
18	31	1.7	55	2	US-08-582-562A-16
19	31	1.7	55	2	US-08-778-494B-16
20	31	1.7	55	3	US-09-294-923-18
21	31	1.7	60	1	US-08-241-465B-11
22	31	1.7	165	2	US-08-783-395-3
23	31	1.7	165	2	US-08-924-83B-9
24	31	1.7	471	5	PCT-US95-13658-1
25	31	1.7	482	4	US-09-843-472-1
26	31	1.7	855	6	5185441-40
27	31	1.7	855	6	5223394-3

28	31	1.7	863	1	US-07-940-861-11	Sequence 11, Appl
29	31	1.7	863	1	US-08-459-512-11	Sequence 9, Appl
30	31	1.7	863	2	US-08-459-512-11	Sequence 11, Appl
31	31	1.7	863	2	US-08-460-132-11	Sequence 11, Appl
32	31	1.7	863	5	PCT-US92-02050-11	Sequence 11, Appl
33	31	1.7	863	6	5185441-35	Patent No. 5185441
34	31	1.7	863	6	5223394-5	Patent No. 5223394
35	31	1.7	1003	2	US-08-887-997B-1	Sequence 1, Appl
36	31	1.7	1024	2	US-09-328-475C-9	Sequence 9, Appl
37	31	1.7	1024	4	US-09-328-475C-20	Sequence 20, Appl
38	31	1.7	1024	4	US-09-328-475C-37	Sequence 37, Appl
39	31	1.7	1338	2	US-08-484-933B-7	Sequence 7, Appl
40	31	1.7	1338	2	US-08-484-158B-7	Sequence 7, Appl
41	31	1.7	1338	2	US-08-484-596A-7	Sequence 7, Appl
42	31	1.7	1338	2	US-08-480-150A-7	Sequence 7, Appl
43	31	1.7	1338	3	US-08-458-731-7	Sequence 7, Appl
44	31	1.7	1338	3	US-08-149-223A-7	Sequence 7, Appl
45	31	1.7	1502	2	US-08-651-940-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
APPLICANT: Zeng, Wenlin
TITLE OF INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/120,646
PRIOR FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8)...(1418)
OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-5

Query Match 71.9%; Score 1336; DB 4; Length 1438;
Best local similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	217	CCCCAGATGGGAAACTTCACTCAAGCCGCGCTGTGAAGCGGAGAGAGCC	276
DB	1	CCCCAGATGGGAAACTTCACTCAAGCCGCGCTGTGAAGCGGAGAGAGCC	60
QY	277	GGAAGTGAAGCTTGCCTGAGCCCTGAGCTCGAAGGAGATCAGAGTGGAT	336
DB	61	GGAAGTGAAGCTTGCCTGAGCCCTGAGCTCGAAGGAGATCAGAGTGGAT	120
QY	337	CGGAGACAGCGCTGCCGCGGTCTTCGGGACCCGACAGCTCGGTCGGGAC	396
DB	121	CGGAGACAGCGCTGCCGCGGTCTTCGGGACCCGACAGCTCGGTCGGGAC	180
QY	397	CATAGGCCAAGAGACCCGAGAGCTCGTGGCGACGTTGAGACACGCTCAGAGGA	456
DB	181	CATAGGCCAAGAGACCCGAGAGCTCGTGGCGACGTTGAGACACGCTCAGAGGA	240
QY	457	AGAGAGGAGAGATTTGGCTGGAATGGCCCTGCTGAGAGACTGACGGCTGGG	516
DB	241	AGAGAGGAGAGATTTGGCTGGAATGGCCCTGCTGAGAGACTGACGGCTGGG	300

OY	511	TAGGGGAGATAGAAAGAAAGTGGAGAGTGAAGGAAACCCGGCCAGGCTCCAAAGAAACA	576
Db	301	CAGCGGAGATAGAAAGAAAGTGGAGAGTGAAGCAACCTCGCCAGGCTCCAAAGAAACA	360
OY	577	GCTGAAGTTTAAAGAGCTCCAGTGGAGAGTGTCTCATGAGAGAGACAGCCGGCAGAGTG	636
Db	361	GCTGAAGTTTAAAGAGCTCCAGTGGAGAGTGTCTCATGAGAGAGACAGCCGGCAGAGTG	420
OY	637	GACCTTCAACCTGTATGACTTTTGAACAACAGGCAAGGTCAACCCGAGAGACATTCACAG	696
Db	421	GACCTTCAACCTGTATGACTTTTGAACAACAGGCAAGGTCAACCCGAGAGACATTCACAG	480
OY	697	CTTGCTGACACACTTATGAGGTGTGGAATCTCTGTGTCAACAATCCCCAATCCAG	756
Db	481	CTTGCTGACACACTTATGAGGTGTGGAATCTCTGTGTCAACAATCCCCAATCCAG	540
OY	757	CAAGATGTGGGGGTAAGTCAACCGTGGCCCCGATGGCAGCCAGACCAAGAGAGAGT	816
Db	541	CAAGATGTGGGGGTAAGTCAACCGTGGCCCCGATGGCAGCCAGACCAAGAGAGAGT	600
OY	817	CTTGTCAATCAGGCTGAACCTGCAGAGGCGAAGGCCCCGAGCAGAGCAAGCCCATGA	876
Db	601	CCTTGTCATCAGGCTGAACCTGCAGAGGCGAAGGCCCCGAGCAGAGCAAGCCCATGA	660
OY	877	GGACTCTGGAGCTGTGGGAAAGAGACAGGAGCCCGCTCAGGTTTCCAGGGTGAACGGC	936
Db	661	GGACTCTGGAGCTGTGGGAAAGAGACAGGAGCCCGCTCAGGTTTCCAGGGTGAACGGC	720
OY	937	CCTGAGCAGTCTGGCTCTACACACATTTGGTATAGAAACATTCGAGAGAGAAACA	996
Db	721	CCTGAGCAGTCTGGCTCTACACACATTTGGTATAGAAACATTCGAGAGAGAAACA	780
OY	997	CTACTTAGATCTCGCCGGGATAGAAAAGTCAACAGTCCCAATTTGGGCTCGCTCCCTTC	1056
Db	781	CTACTTAGATCTCGCCGGGATAGAAAAGTCAACAGTCCCAATTTGGGCTCGCTCCCTTC	840
OY	1057	CGTGGCCCCAGAAAGTGAACATGCCCCCGGCACCTCCAAATCCCACTCGATCTCGTCCCA	1116
Db	841	CGTGGCCCCAGAAAGTGAACATGCCCCCGGCACCTCCAAATCCCACTCGATCTCGTCCCA	900
OY	1117	TGAGCCGGAAGCAATCCATCCCAACCGAAAGCCCCAGAGCGTGAACCGGACCTCCTT	1176
Db	901	TGAGCCGGAAGCAATCCATCCCAACCGAAAGCCCCAGAGCGTGAACCGGACCTCCTT	960
OY	1177	CCACTTCTTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTTCGGGGCAC	1236
Db	961	CCACTTCTTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTTCGGGGCAC	1020
OY	1237	CCAGGACGGGAGCAAGCACTTTGTATAGTCCCCCAAGGCCAGAGGCAAGATGTGGTGT	1296
Db	1021	TCAGGACGGGAGCAAGCACTTTGTATAGTCCCCCAAGGCCAGAGGCAAGATGTGGTGT	1080
OY	1297	GGGCGCACGTGGCGAGAGGGGGCAAGAAACAAGCCCCCTGTGGGACCCGCGCATCCCGGCGT	1356
Db	1081	GGGCGCACGTGGCGAGAGGGGGCAAGAAACAAGCCCCCTGTGGGACCCGCGCATCCCGGCGT	1140
OY	1357	GTCCCCCTCGGCCCACTGTGCTGACAGCCCGGCTCTCTCCCTTCTCTAGGCCCTTGG	1416
Db	1141	GTCCCCCTCGGCCCACTGTGCTGACAGCCCGGCTCTCTCCCTTCTCTAGGCCCTTGG	1200
OY	1417	GCACAAAGAACCAAGACACCGAGCCCAAGAGACCAAGAGGCTGACGAGGCTTCAGAGC	1476
Db	1201	GCACAAAGAACCAAGACACCGAGCCCAAGAGACCAAGAGGCTGACGAGGCTTCAGAGC	1260
OY	1477	AACCATGTGGCTCAGGTTGGCCCTGTCTCTGTGGGCGGGAGACACTTGGGGGAGCTGCCGCTT	1536
Db	1261	AACCATGTGGCTCAGGTTGGCCCTGTCTCTGTGGGCGGGAGACACTTGGGGGAGCTGCCGCTT	1320
OY	1537	GGTGTGTATAGAGCCCAAGCGCGGAGACCGGTTCAGAGACATGAGCACACACACCA	1596
Db	1321	GGTGTGTATAGAGCCCAAGCGCGGAGACCGGTTCAGAGACATGAGCACACACACCA	1380
OY	1597	TGAACATCACCACTTACACACTTCTTACCAAGCATAGAGCCCTTCCAGGGCC	1654

```

Db      1381 TGAACATCACCACTTACCACTTCTACCAAGCATAGAGCCCTCCCGAGGGCC 1438
RESULT 2
US-10-101-464A-949
; Sequence 949, Application US/10101464A
; Patent No. 6768041
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 949
; LENGTH: 3768
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-101-464A-949

Query Match      2.9%; Score 53; DB 4; Length 3768;
Best Local Similarity 100.0%; Pred. No. 7.1e-12;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      13 CTAATACGACTCATCTATAGGGGCAAGCAGTGTAAACAACCAAGTACGGGGG 65
      1 CTAATACGACTCATCTATAGGGGCAAGCAGTGTAAACAACCAAGTACGGGGG 53

RESULT 3
US-09-506-066E-3
; Sequence 3, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)..(1553)
; OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-3

Query Match      2.7%; Score 50; DB 4; Length 1731;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      242 AAGCCGGCCGCGTGTGCAAGCGCAGGAGAGACGCCGGAAGGTACAGACTT 251

```

Db 158 AAGCCGCGCGCTGTGCAAGCGAGAGAGCCCGAAGGTACAGACTT 207

RESULT 4

US-09-686-583B-65
Sequence 65, Application US/09666583B
Patent No. 6576750
GENERAL INFORMATION:
APPLICANT: Heeska Corporation
APPLICANT: Gaines, Patrick J.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: FLUA PERITROPHIN NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
FILE REFERENCE: FC-6-C2
CURRENT APPLICATION NUMBER: US/09/686,583B
CURRENT FILING DATE: 2000-10-11
PRIOR APPLICATION NUMBER: 09/543,668
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 60/128,704
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 68
SOFTWARE: PatentIn version 3.1
SEQ ID NO 65
LENGTH: 45
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Synthetic Primer
US-09-686-583B-65

Query Match

2.4%; Score 45; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 CTATAGCACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 57
Db 1 CTATAGCACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 45

RESULT 5

US-09-672-725C-18
Sequence 18, Application US/09672725C
Patent No. 6753177
GENERAL INFORMATION:
APPLICANT: Stocker, Penny J.
APPLICANT: Steimel-Crespi, Dorothy T.
APPLICANT: Crespi, Charles L.
APPLICANT: Rief, Timothy C.
TITLE OF INVENTION: P-GLYCOPROTEINS AND USES THEREOF
FILE REFERENCE: G0307/7017
CURRENT APPLICATION NUMBER: US/09/672,725C
CURRENT FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/156,510
PRIOR FILING DATE: 1999-09-28
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 18
LENGTH: 45
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide
US-09-672-725C-18

Query Match

2.4%; Score 45; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 CTATAGCACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 57
Db 1 CTATAGCACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 45

RESULT 6
US-08-778-494B-114/c
Sequence 114, Application US/08778494B
Patent No. 5962272

GENERAL INFORMATION:

APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
TITLE OF INVENTION: CLONING
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESS: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,494B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/582,562
FILING DATE: 03-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Dorian R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: CL-7C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800

INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-114

Query Match 1.7%; Score 31; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAAGCGCGCGC 1859
Db 38 AAAAAAAAAAAAAAAAAAAGCGCGCGC 8

RESULT 7

US-08-381-572-20/c
Sequence 20, Application US/08381572
Patent No. 5565340

GENERAL INFORMATION:

APPLICANT: Chenchik, Alex
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
APPLICANT: Lukianov, Sergey
APPLICANT: Gurekaya, Nadia
APPLICANT: Tarabukin, Victor
APPLICANT: Sverdlov, Eugene
TITLE OF INVENTION: METHOD FOR SUPPRESSING DNA FRAGMENT
TITLE OF INVENTION: AMPLIFICATION DURING PCR
NUMBER OF SEQUENCES: 26

```

CORRESPONDENCE ADDRESS:
ADDRESS: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,572
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,361
REFERENCE/DOCKET NUMBER: CL-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-381-572-20

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
Db 43 AAAAAAAAAAAAAAAAAAGCGCGCGC 13

RESULT 8
US-08-592-820-20/c
Sequence 20, Application US/08592820
Patent No. 5759822
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
APPLICANT: Lukianov, Sergey
APPLICANT: Gurskaya, Nadia
APPLICANT: Tarebkyin, Victor
APPLICANT: Sverdlov, Eugene
TITLE OF INVENTION: METHOD FOR SUPPRESSING DNA FRAGMENT
TITLE OF INVENTION: AMPLIFICATION DURING PCR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESS: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,820
FILING DATE:
CLASSIFICATION: 435
```

```

ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,361
REFERENCE/DOCKET NUMBER: CL-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-592-820-20

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
Db 43 AAAAAAAAAAAAAAAAAAGCGCGCGC 13

RESULT 9
US-09-129-603-3/c
Sequence 3, Application US/09129603A
Patent No. 6790944
GENERAL INFORMATION:
APPLICANT: Ishiwata, Tetsuyoshi
APPLICANT: Sakurada, Mikiko
APPLICANT: Nishimura, Ayako
APPLICANT: Nakagawa, Satoshi
APPLICANT: Kuga, Tetsuro
APPLICANT: Nishi, Tatsunari
APPLICANT: No. 6790944ura, No. 6790944uo
APPLICANT: Sawada, Shigemasa
APPLICANT: Nagase, Takahiro
APPLICANT: Takei, Masami
TITLE OF INVENTION: No. 6790944el Protein
FILE REFERENCE: 766.25
CURRENT APPLICATION NUMBER: US/09/129,603A
CURRENT FILING DATE: 1998-08-05
EARLIER APPLICATION NUMBER: PCT/JP97/04469
EARLIER FILING DATE: 1997-12-05
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 50
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: other nucleic acid from homo sapiens, synthesized
US-09-129-603-3

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
Db 39 AAAAAAAAAAAAAAAAAAGCGCGCGC 9

RESULT 10
US-08-582-562A-8/c
Sequence 8, Application US/08582562A
Patent No. 5962271
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
```

APPLICANT: Diachenko, Lida
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHOD FOR FULL-LENGTH CDNA CLONING
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/582,562A
FILING DATE: 03-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: CL-7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
LENGTH: 51 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-582-562A-8

Query Match 1.7%; Score 31; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
Db 42 AAAAAAAAAAAAAAAAAAGCGCGCGC 12

RESULT 11
US-08-778-494B-8/c
Sequence 8, Application US/08778494B
Patent No. 5962272
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
APPLICANT: Diachenko, Lida
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
TITLE OF INVENTION: CLONING
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,494B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/582,562
FILING DATE: 03-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: CL-7C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-8

Query Match 1.7%; Score 31; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
Db 42 AAAAAAAAAAAAAAAAAAGCGCGCGC 12

RESULT 12
US-08-859-998-1373/c
Sequence 1373, Application US/08859998
Patent No. 5994076
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Jorhadze, George
APPLICANT: Bibilashvili, Robert
TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 1375
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/859,998
FILING DATE: 21-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1373:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-859-998-1373

Query Match 1.7%; Score 31; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAGCGCCGC 1859
42 AAAAAAAAAAAAAAAAAAGCGCCGC 12

Db

RESULT 13
US-09-225-928-1373/C
Sequence 1373, Application US/09225928
Patent No. 6352829
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
Jokhadze, George
Bibilashvili, Robert
TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
EXPRESSION
NUMBER OF SEQUENCES: 1375
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225.928
FILING DATE: 05-Jan-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/859,998
FILING DATE: 21-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1373:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1373:
US-09-225-928-1373

Query Match 1.7%; Score 31; DB 3; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAGCGCCGC 1859
42 AAAAAAAAAAAAAAAAAAGCGCCGC 12

Db

RESULT 14
US-09-225-201B-1373/C
Sequence 1373, Application US/09225201B
Patent No. 6489455
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
Jokhadze, George

Bibilashvili, Robert
TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
EXPRESSION
NUMBER OF SEQUENCES: 1375
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225.201B
FILING DATE: 05-Jan-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/859,998
FILING DATE: 21-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1373:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1373:
US-09-225-201B-1373

Query Match 1.7%; Score 31; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAGCGCCGC 1859
42 AAAAAAAAAAAAAAAAAAGCGCCGC 12

Db

RESULT 15
US-08-618-100B-9/C
Sequence 9, Application US/08618100B
Patent No. 6068976
GENERAL INFORMATION:
APPLICANT: Briggs, Michael R.
APPLICANT: Auwerx, Johan
APPLICANT: de Vos, Piet
APPLICANT: Steels, Bart
APPLICANT: Croston, Glenn E.
APPLICANT: Miller, Stephen G.
TITLE OF INVENTION: MODULATORS OF OB GENE AND
TITLE OF INVENTION: SCREENING METHODS THEREFOR
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

```

MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,100B
FILING DATE: March 19, 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/558,588
FILING DATE: October 30, 1995
APPLICATION NUMBER: 08/510,584
FILING DATE: August 2, 1995
APPLICATION NUMBER: 08/418,096
FILING DATE: April 5, 1995
APPLICATION NUMBER: 08/408,584
FILING DATE: March 20, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 219/075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: "N" represents any base.
US-08-618-100B-9

Query Match      1.7% Score 31; DB 3; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1829 AAAAAAAAAAAAAAAAAAGCGGCCGC 1859
DB      43 AAAAAAAAAAAAAAAAAAGCGGCCGC 13

Search completed: December 30, 2004, 03:23:40
Job time : 213 secs
```

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 20:34:24 ; Search time 8793 Seconds
(without alignments)
7704.022 Million cell updates/sec

Title: US-09-993-966-5

Sequence: 1 gaattcgcccttctaatacgcg.....aaaaaaaaaagcgcgcgc 1859

Scoring table: (OLIGO NUC)
Gapop 60.0, Gapext 60.0

Searched: 32822875 seqs, 18219865908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1525	82.0	2142	3 AF289584	AF289584 Homo sapi
2	885	47.6	885	9 AY412098	AY412098 Homo sapi
3	814	43.8	931	5 B0653673	B0653673 AGENCOURT
4	748	40.2	921	5 B0645656	B0645656 AGENCOURT
5	702	37.8	953	5 B0846054	B0846054 AGENCOURT
6	689	37.1	990	5 B0064678	B0064678 AGENCOURT
7	660	35.5	895	5 B0644360	B0644360 AGENCOURT
8	659	35.4	928	5 B0644956	B0644956 AGENCOURT
9	640	34.4	996	5 B0652087	B0652087 AGENCOURT
10	639	34.4	966	5 B0649813	B0649813 AGENCOURT
11	622	33.5	622	6 CB215756	CB215756 NISC np09
12	610	32.8	825	6 CB961961	CB961961 AGENCOURT
13	593	31.9	634	7 CN369119	CN369119 170005999
14	582	31.3	627	4 B1767278	B1767278 603057995
15	574	30.9	1027	4 B0104777	B0104777 602311726
16	556	29.9	1008	5 B0646371	B0646371 AGENCOURT
17	551	29.6	826	6 CB961668	CB961668 AGENCOURT
18	539	29.0	1059	5 B0070932	B0070932 AGENCOURT
19	508	27.3	364	4 BM711145	BM711145 UT-E-DX1-
20	465	25.0	899	5 B0645507	B0645507 AGENCOURT
21	452	24.3	885	9 AY412099	AY412099 Pan trogl
22	450	24.2	751	4 BG820139	BG820139 602782356
23	446	24.0	931	5 B0856404	B0856404 AGENCOURT
24	429	23.1	481	4 B1047069	B1047069 MRS-FN020

25	397	21.4	777	4 BG542261	BG542261 602571809
26	395	21.2	395	1 A1167910	A1167910 OKR9B05.X
27	334	18.0	429	7 CN369118	CN369118 170005322
28	326	17.5	654	9 AG112447	AG112447 Pan trogl
29	244	13.1	333	2 BF436193	BF436193 nab45a02.
30	220	11.8	890	5 B0645220	B0645220 AGENCOURT
31	220	11.8	959	5 B0653014	B0653014 AGENCOURT
32	209	11.2	555	2 BF920913	BF920913 MR2-NT013
33	188	10.1	500	4 BM670652	BM670652 UT-E-DX1-
34	126	6.8	638	9 AG053170	AG053170 Pan trogl
35	104	5.6	550	2 AW752389	AW752389 RC2-CT020
36	78	4.2	555	5 BP446790	BP446790 BP446790
37	70	3.8	684	7 CK450648	CK450648 903739 MA
38	61	3.3	252	8 B91703	B91703 RBC111-20A1
39	60	3.2	1487	6 CB381772	CB381772 UZ1T1L5.2
40	56	3.0	661	1 AU260449	AU260449 AU260449
41	55	3.0	631	5 B0622880	B0622880 ZOCR24 C
42	54	2.9	214	5 B0622887	B0622887 ZOCR24 C
43	54	2.9	254	5 B0622885	B0622885 ZOCR24 C
44	54	2.9	639	5 B0622909	B0622909 ZOCR24 C
45	54	2.9	707	5 B0622883	B0622883 ZOCR24 C

ALIGNMENTS

RESULT 1	AF289584	2142 bp	MRNA	linear	HTC 01-JAN-2002
LOCUS	AF289584		Homo sapiens clone p7246	unknown	MRNA.
DEFINITION	AF289584				
ACCESSION	AF289584.1	GI:18027371			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
CDS					
ORIGIN					
Query Match	82.0%	Score 1525	DB 3	Length 2142	

```
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1575; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

OY	253	CGTGTGCAAAGCGCAGGAGGAGCCCGGAAAGTGAACAAGCTTCCCGTGAACGCGCTGCTGGGAC	312
Db	306	CGTGTGCAAAGCGCAGGAGAGAGCCCGGAAAGTGAACAAGCTTCCCGTGAACGCGCTGCTGGGAC	365
OY	313	TCGGAAAGGCAATCGAGAGTGGATTCGGGAGACAGCGCTGCCCGGCGGTGTCTCGGAGCC	372
Db	366	TCGGAAAGGCAATCGAGAGTGGATTCGGGAGACAGCGCTGCCCGGCGGTGTCTCGGAGCC	425
OY	373	CCGACAGCTGGGTTCGGGGGCAACAATAGCCGGAACCCGGAGCTTCGTGGGCGACGT	432
Db	426	CCGACAGCTGGGTTCGGGGGCAACAATAGCCGGAACCCGGAGCTTCGTGGGCGACGT	485
OY	433	GTTTGAGAACACGCTCAACGAGGAAGAGAGAGACATTTGGGCTGGAAAGTGGCCCTTCC	492
Db	486	GTTTGAGAACACGCTCAACGAGGAAGAGAGAGACATTTGGGCTGGAAAGTGGCCCTTCC	545
OY	493	TCCTGAGAAAGCTGACGGGCTTGGGACGCGAGATGATGAGAAAGATGAGAGAGTGAAGCA	552
Db	546	TCCTGAGAAAGCTGACGGGCTTGGGACGCGAGATGATGAGAAAGATGAGAGAGTGAAGCA	605
OY	553	ACCCTGCCAGGCTTCAAGAAAGCAGCTGAAGTTGAAGAGCTCCAGTGCAGCGTGTCCAT	612
Db	606	ACCCTGCCAGGCTTCAAGAAAGCAGCTGAAGTTGAAGAGCTCCAGTGCAGCGTGTCCAT	665
OY	613	GGAGGAGGACAGCCGGGCGAGGTGGACCTTCAACCTGTATATACATTTTGAACAACAACGGCA	672
Db	666	GGAGGAGGACAGCCGGGCGAGGTGGACCTTCAACCTGTATATACATTTTGAACAACAACGGCA	725
OY	673	GGTCAACCCGAGAGGACATCAACAGCTTTCGTGCACACATCTATGAGTGGTGGACTCTCTC	732
Db	726	GGTCAACCCGAGAGGACATCAACAGCTTTCGTGCACACATCTATGAGTGGTGGACTCTCTC	785
OY	733	TGTCAACCACTTCCCCAATATCCAGCAGATGCTGCGGGTAAAGCTCACCGTGGCCCCCGA	792
Db	786	TGTCAACCACTTCCCCAATATCCAGCAGATGCTGCGGGTAAAGCTCACCGTGGCCCCCGA	845
OY	793	TGGCAGCCGAGAGCAAGAGAGGGTCTCTGTCAATTAGGCTGACCTGCGAGACCGCAAGGCC	852
Db	846	TGGCAGCCGAGAGCAAGAGAGGGTCTCTGTCAATTAGGCTGACCTGCGAGACCGCAAGGCC	905
OY	853	CCGAGCAGAGACCAAGCCCACTGAGAGACCTGCGGAGCTGGGAGAGAAAGACAGCAGCCCC	912
Db	906	CCGAGCAGAGACCAAGCCCACTGAGAGACCTGCGGAGCTGGGAGAGAAAGACAGCAGCCCC	965
OY	913	GCTCAGGTTCAAGGGTGAACAGCCGCTTGGAGCAAGTCTGGCTGTACCAACAATTGGTGA	972
Db	966	GCTCAGGTTCAAGGGTGAACAGCCGCTTGGAGCAAGTCTGGCTGTACCAACAATTGGTGA	1025
OY	973	TGAGAACATGAGAGGAGAAACAATACTTATGATTCGCCCGGATATGAAAACTACACGTC	1032
Db	1026	TGAGAACATGAGAGGAGAAACAATACTTATGATTCGCCCGGATATGAAAACTACACGTC	1085
OY	1033	CCAAATTTGGGCTGGCTCCCTTCGCTGGGCCACAGATGCAGAACTGCCCCCGCAGACTC	1092
Db	1086	CCAAATTTGGGCTGGCTCCCTTCGCTGGGCCACAGATGCAGAACTGCCCCCGCAGACTC	1145
OY	1093	CAATCCACTCGATTCGCTCCCATAGACCGGAAACCATCAATCCCAACCGAAAGCC	1152
Db	1146	CAATCCACTCGATTCGCTCCCATAGACCGGAAACCATCAATCCCAACCGAAAGCC	1205
OY	1153	CCAAAGGCGTGAACCCGGGCTCTTCCATTCCTTGAACCCCAATTCGCAAGTCTCAGA	1212
Db	1206	CCAAAGGCGTGAACCCGGGCTCTTCCATTCCTTGAACCCCAATTCGCAAGTCTCAGA	1265
OY	1213	GCTCCAGCAACGAGCTCCGGGGCACCCAGACGGGAGCAAGCACTTGTGAGGTCCCCCA	1272
Db	1266	GCTCCAGCAACGAGCTCCGGGGCACCCAGACGGGAGCAAGCACTTGTGAGGTCCCCCA	1325
OY	1273	GGCCCAAGGCAAGAGTGTGGGTGTGGGCAACGTGGCCAGAGGGGCAAGAAACAAGCCCC	1332

Db	1326	GGCCACAGGAGCAAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAACAAGCCCC	1385
QY	1333	TCTGGAGACCCGCACATCCCTGGCGATGTCCCTCCGCCACCTGGCTGCGAGCCGGCCTT	1392
Db	1386	TCTGGGACCCGGCCATCCCTGGCGATGTCCCTCCGCCACCTGGCTGCGAGCCGGCCTT	1445
QY	1393	CCTCCCTCTCTTAAGCCCCCTCTGGGACACAGAACACAGATCCGAGCCAAAGAGAGCCA	1452
Db	1446	CCTCCCTCTCTTAAGCCCCCTCTGGGACACAGAACACAGATCCGAGCCAAAGAGAGCCA	1505
QY	1453	GAAGGGCTGCGGGGGCCTGACAGGACACATGGGCTCAGGTGGCCCTGTCCTGGGGCGGGA	1512
Db	1506	GAAGGGCTGCGGGGGCCCTGACAGGACACATGGGCTCAGGTGGCCCTGTCCTGGGGCGGGA	1565
QY	1513	GCACCTGCGGGAGCTGCCCCGCTTGTTGTGTGTATAGAGGCCAGGCGGGACGCCGTCCA	1572
Db	1566	GCACCTGCGGGAGCTGCCCCGCTTGTTGTGTGTATAGAGGCCAGGCGGGAGCGCGGTCCA	1625
QY	1573	GAGACATGAGCACACACACACATGAACATCACCAACATTAACACACATTCTTACAGAC	1632
Db	1626	GAGACATGAGCACACACACACATGAACATCACACACATTAACACACATTCTTACAGAC	1685
QY	1633	ATAGAGCCCCCTCCCCAGGGCCCCCACCCTGCATATGAAGAGCCCCACCCCGAATACACA	1692
Db	1686	ATAGAGCCCCCTCCCCAGGGCCCCCACCCTGCATATGAAGAGCCCCACCCCGAATACACA	1745
QY	1693	AGGCATTATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATG	1752
Db	1746	AGGCATTATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATG	1805
QY	1753	TTACTCCACTAATATTTAGCTAGCTTACATGTAGTAAGATCTTAGAACAACAGAACTPAA	1812
Db	1806	TTACTCCACTAATATTTAGCTAGCTTACATGTAGTAAGATCTTAGAACAACAGAACTPAA	1865
QY	1813	CTTTTATTTATATGTT 1828	
Db	1866	CTTTTATTTATATGTT 1881	

[illegible]

QY 1182 TCCTTGACACCCCAATGCGCAAGGCTTCAGAGCTCCAGCAAGGCTCCGGGGCAACCGAG 1241
DB 361 TCTTGTGACACCCCAATGCGCAAGGCTTCAGAGCTCCAGCAAGGCTCCGGGGCAACCGAG 420
QY 1242 ACCGAGCAAGCACTTTGTAGAGTCCGCCAAGGCGCGAGCAAGTGTGGGTGGGCC 1301
DB 421 ACCGAGCAAGCACTTTGTAGAGTCCGCCAAGGCGCGAGCAAGTGTGGGTGGGCC 480
QY 1302 ACGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCCATCTCTGGGTGCC 1361
DB 481 ACGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCCATCTCTGGGTGCC 540
QY 1362 CCTCCGCCACCTGGCTGGCCAGCGCGGCTCTCCCTCCCTTACCCCTCCCTGGGGACA 1421
DB 541 CCTCCGCCACCTGGCTGGCCAGCGCGGCTCTCCCTCCCTTACCCCTCCCTGGGGACA 600
QY 1422 AGAAGCAACAAGCAAGGCGCAAGGAGAGCCAGAGGCTGCGGGGCTGAGGCAACAC 1481
DB 601 AGAAGCAACAAGCAAGGCGCAAGGAGAGCCAGAGGCTGCGGGGCTGAGGCAACAC 660
QY 1482 TGGCTCAGGTGGCTCTGTCTTGGGGGCGGAGCACTTGGGAGCTGCCCTTTGGTG 1541
DB 661 TGGCTCAGGTGGCTCTGTCTTGGGGGCGGAGCACTTGGGAGCTGCCCTTTGGTG 720
QY 1542 TGTATGAGAGCGAGGCGGGGCGGCTCCAGAGCAATGAGCAACACCAACCATGAAAC 1601
DB 721 TGTATGAGAGCGAGGCGGGGCGGCTCCAGAGCAATGAGCAACACCAACCATGAAAC 780
QY 1602 ATCACCACTATTACCACTTCTTACCAAGACATA 1635
DB 781 ATCACCACTATTACCACTTCTTACCAAGACATA 814
RESULT 4
LOCUS B0645656 921 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT 8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
5', mRNA sequence.
ACCESSION B0645656
VERSION B0645656.1 GI:21769828
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 921)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: rgapbs-rc@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LNCM2484 row: d column: 05
High quality sequence start: 16
High quality sequence stop: 685.
FEATURES
location/Qualifiers
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6285268"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_100"
/note="Organ: liver; Vector: POT87; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned"

intc EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using Zap-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Query Match 40.2%; Score 748; DB 5; Length 921;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 748; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 ATCAGCTGACCTGCGAGAGGCAAGGCGCGAGCAAGACCAAGCCACTGAGGACCTGC 884
DB 23 ATCAGCTGACCTGCGAGAGGCAAGGCGCGAGCAAGACCAAGCCACTGAGGACCTGC 82
QY 885 GAGCTGGAGAAAGAAAGCAAGCGCCCGCTCAGGTTCCAGGGTGAAGCGCGCTGGAGC 944
DB 83 GAGCTGGAGAAAGAAAGCAAGCGCCCGCTCAGGTTCCAGGGTGAAGCGCGCTGGAGC 142
QY 945 AGTCTGCTCTACCACTTGGCTAGATGAGAACTCAGAGAGAAACCACTACTTAG 1004
DB 143 AGTCTGCTCTACCACTTGGCTAGATGAGAACTCAGAGAGAAACCACTACTTAG 202
QY 1005 ATCTCGCGGGATAGAAAATACACGTCCTCAATTTGGGCTCTGCTCCCTTCCGTGAGCC 1064
DB 203 ATCTCGCGGGATAGAAAATACACGTCCTCAATTTGGGCTCTGCTCCCTTCCGTGAGCC 262
QY 1065 AGAAGTCAGAACTGCGCCCGCGGACCTCCATCCACTGATCTGCTCCCATGAGCGG 1124
DB 263 AGAAGTCAGAACTGCGCCCGCGGACCTCCATCCACTGATCTGCTCCCATGAGCGG 322
QY 1125 AAGCCATCCATATCCCAACCGAAAGCCCAAGGCTGACCCGGCTCTTCCACTTCC 1184
DB 323 AAGCCATCCATATCCCAACCGAAAGCCCAAGGCTGACCCGGCTCTTCCACTTCC 382
QY 1185 TTGACACCCCAATTCGCAAGGCTCTCAGAGCTCCAGCAACGAGCTCCGGGAGCAAGAGC 1244
DB 383 TTGACACCCCAATTCGCAAGGCTCTCAGAGCTCCAGCAACGAGCTCCGGGAGCAAGAGC 442
QY 1245 GAGCAAGCACTTGTGAGGTCCCGCAAGGCGCGAGGCAAGAGTGTGGTGGCGCAGC 1304
DB 443 GAGCAAGCACTTGTGAGGTCCCGCAAGGCGCGAGGCAAGAGTGTGGTGGCGCAGC 502
QY 1305 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCCGCATTCCTGCGGTGCCCT 1364
DB 503 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCCGCATTCCTGCGGTGCCCT 562
QY 1365 CCGGCCACCTGGCTGCCAGGCGCGGCTCTCCCTCCCTTACGCCCCCTCGGGGCAAGA 1424
DB 563 CCGGCCACCTGGCTGCCAGGCGCGGCTCTCCCTCCCTTACGCCCCCTCGGGGCAAGA 622
QY 1425 AGCAACAAGCAAGCAAGCAAGAGAGCAAGGAGGCTGCGGGGCTGAGGCAACCACTGG 1484
DB 623 AGCAACAAGCAAGCAAGCAAGAGAGCAAGGAGGCTGCGGGGCTGAGGCAACCACTGG 682
QY 1485 CTTGAGGTGGCTCTGTCTTGGGGCGGAGCACTCTGCGGAGCTGCGGCTTGTGGTGT 1544
DB 683 CTTGAGGTGGCTCTGTCTTGGGGCGGAGCACTCTGCGGAGCTGCGGCTTGTGGTGT 742
QY 1545 ATGAGGCGAGGCGGGGAGCGGATCCA 1572
DB 743 ATGAGGCGAGGCGGGGAGCGGATCCA 770
RESULT 5
LOCUS BU846054 953 bp mRNA linear EST 16-OCT-2002
DEFINITION AGENCOURT 10413301 NIH_MGC_109 Homo sapiens cDNA clone
IMAGE:6580002 5', mRNA sequence.
ACCESSION BU846054
VERSION BU846054.1 GI:24030759

```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 953)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strauberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM2784 row: d column: 18
High quality sequence stop: 627.
Location/Qualifiers
1. 953
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6580002"
/tissue_type="teratocarcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 109"
/notes="Organ: ovary; Vector: pOTB7, Site 1: EcoRI, Site 2:
XhoI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC library."

ORIGIN
Query Match 37.8%; Score 702; DB 5; Length 953;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 702; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 AGTCGGGCGCGCGAGCGCGCGAGCGCGCTCCGCGCGCTCGGCTCGG 125
DB 66 AGTCGGGCGCGCGAGCGCGCGAGCGCGCTCCGCGCGCTCGGCTCGG 125
QY 126 CTGCGGGGCGCTTCGGGAGGAGGAGCCAGAGGAGCGCGCGCGCGGCG 165
DB 126 CTGCGGGGCGCTTCGGGAGGAGGAGCCAGAGGAGCGCGCGCGCGGCG 185
QY 186 CATGCTTAAAGAGCGCTCCGCGCGCGAGCGCGCATGAGGAACTTCATCCAAAGC 245
DB 186 CATGCTTAAAGAGCGCTCCGCGCGCGAGCGCGCATGAGGAACTTCATCCAAAGC 245
QY 246 CGGCGCGCGCTGTGCAAGCGAGGAGAGCCCGGAGGTGACAGCTTGCCTGAGCGCTG 305
DB 246 CGGCGCGCGCTGTGCAAGCGAGGAGAGCCCGGAGGTGACAGCTTGCCTGAGCGCTG 305
QY 306 CCGGGGCTCGGAGGAGGATGAGAGAGTGATTCGGGAGAGCGCTGCGCGCGGTCTT 365
DB 306 CCGGGGCTCGGAGGAGGATGAGAGAGTGATTCGGGAGAGCGCTGCGCGCGGTCTT 365
QY 366 CGGAGCCCCGAGAGCTGCGGTGGCGGAGCAATAGCCCAAGACCCGGAAGCTCGTG 425
DB 366 CGGAGCCCCGAGAGCTGCGGTGGCGGAGCAATAGCCCAAGACCCGGAAGCTCGTG 425
QY 426 GCGACGCTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGCTTTCGCTGGAAGTGG 485
DB 426 GCGACGCTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGCTTTCGCTGGAAGTGG 485
QY 486 CCCTGCTCTGGAAGAGCTGACGGGGCTGGGAGCGGAGATGAGAAAGATGAGAGAG 545
DB 486 CCCTGCTCTGGAAGAGCTGACGGGGCTGGGAGCGGAGATGAGAAAGATGAGAGAG 545

FEATURES
source
1. 990
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5926427"
/tissue_type="lymphoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 99"
/notes="Organ: lymph; Vector: pOTB7, Site 1: XhoI, Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the Laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
library."

ORIGIN
Query Match 37.1%; Score 689; DB 5; Length 990;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 739; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 822 TCAATCAGGCTGACCTGACAGAGCGCAAGCGCGAGAGAGCAAGCCCACTGAGAGCC 881
DB 1 TCAATCAGGCTGACCTGACAGAGCGCAAGCGCGCGAGAGAGCAAGCCCACTGAGAGCC 60

```


[illegible]

Db	199	ATCTGCGGGATAGAAAACTACACGCTCCAAATT766GCGTGGCTCCCGTTCGGTGGCC	258
Oy	1065	AGAGTCAAACTGCCCCCCCCGACCTTCGATCCCACTCGATCTCGCTCCATAGCCGG	1124
Db	259	AGAGTCAAACTGCCCCCCCCGACCTTCGATCCCACTCGATCTCGCTCCATAGCCGG	318
Oy	1125	AAGCATTCACATCCCAACCCGAAAGGCCAAGGCGTGAACCCGGGCTCCCTCCACTTC	1184
Db	319	AAGCATTCACATCCCAACCCGAAAGGCCAAGGCGTGAACCCGGGCTCCCTCCACTTC	378
Oy	1185	TTGACACCCCAATCGCCCAAGGCTCAGAGCTCCAGCAACGGCTCCGGGAGCAACAGAGACG	1244
Db	379	TTGACACCCCAATCGCCCAAGGCTCAGAGCTCCAGCAACGGCTCCGGGAGCAACAGAGACG	438
Oy	1245	GGAGCAAGCACTTTGTAGAGTCCCCCAAGGCCCAAGGCGCAAGAGTGTGGGTGGCAACG	1304
Db	439	GGAGCAAGCACTTTGTAGAGTCCCCCAAGGCCCAAGGCGCAAGAGTGTGGGTGGCAACG	498
Oy	1305	TGGCCAAAGGGGGCAAGAAACAAGCCCCCTCGTGGGACCCGCGATCCCGCGGTCCCTT	1364
Db	499	TGGCCAAAGGGGGCAAGAAACAAGCCCCCTCGTGGGACCCGCGATCCCGCGGTCCCTT	558
Oy	1365	CCGCCCACTTGTGGTGCAGCCCGGCGCTCTCTCCCTCTAGACCCCTCGGAGCAAGA	1424
Db	559	CCGCCCACTTGTGGTGCAGCCCGGCGCTCTCTCCCTCTAGACCCCTCGGAGCAAGA	618
Oy	1425	AGCAACAAGCAACGAGCCAAAGAGAGCAGCAGGAGCTCCGGGAGCTCAGGCACTAGG	1484
Db	619	AGCAACAAGCAACGAGCCAAAGAGAGCAGCAGGAGCTCCGGGAGCTCAGGCACTAGG	678
Oy	1485	CCTCAGGTGGCTCTCTCTGGGGGGGAGCACTTCGGGAGCTGGCCCGCC	1544
Db	679	CCTCAGGTGGCTCTCTCTGGGGGGGAGCACTTCGGGAGCTGGCCCGCC	728
RESULT 9			
LOCUS	B0652087	996 bp	linear
DEFINITION	AGENCOURT 8207495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283110		EST 15-JUL-2002
ACCESSION	B0652087		
VERSION	B0652087.1	GI:21776259	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	NIH-MGC http://mgs.nci.nih.gov/.		
JOURNAL	1 (bases 1 to 996)		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@db-jemmail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LUCM2478 row: 1 column: 07 High quality sequence stop: 604. Location/Qualifiers 1..996 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6283110" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_id="NIH_MGC_100" /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:		

Qy	1305	TGGCGAAGGGGGGCAAGAAACAAAGCCCTCTGTGGGACCGGCATCTCGTGGGTTCCCT	1364
Db	483	TGGCGAAGGGGGGCAAGAAACAAAGCCCTCTGTGGGACCGGCATCTCTCGGTTCCCT	542
Qy	1365	CCGGCCACTTGTGGCTGCGCAGCCCGGCTCTCTCCCTCCCTAGCCCTCCGCGGCAAGA	1424
Db	543	CCGGCCACTTGTGGCTGCGCAGCCCGGCTCTCTCCCTCCCTAGCCCTCCGCGGCAAGA	602
Qy	1425	AGCACAAGCACCAGACCAGAGAGACAGAGGCTGCGGCGCTCGAGGACCACTGG	1484
Db	603	AGCACAAGCACCAGACCAGAGAGAGACAGAGGCTGCGGCGCTCGAGGACCACTGG	662
Qy	1485	CCTCAGGTGGCCCTGTCTCGGGGCGGGAGC	1514
Db	663	CCTCAGGTGGCCCTGTCTCGGGGCGGGAGC	692
RESULT 11			
LOCUS	CB215756		
DEFINITION	CB215756	622 bp	mRNA linear EST 06-FEB-2003
ACCESSION	NISC_np09612.y1 NICHDS_HS_Ut1	Homo sapiens	CDNA IMAGE:5937502
VERSION	CB215756		
KEYWORDS	CB215756.1	GI:28263948	
ORGANISM	EST..		
REFERENCE	Homo sapiens (human)		
AUTHORS	Homo sapiens		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
UNPUBLISHED	Unpublished (1997)		
CONTACT	Contact: Robert Strausberg, Ph.D.		
EMAIL	Email: cga@bbs-rcmail.nih.gov		
CDNA	CDNA Library Preparation:		
DNA	DNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL		
SEQUENCING	CDNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)		
CLONE	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: info@image.jhmi.gov		
PLATE	Plate: LLM13164	row: I	column: 23
SEQ	Seq primer: M13RPI reverse primer (ABI).		
LOCATION	Location/Qualifiers		
1..622			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5937502"			
/sex="female"			
/tissue_type="normal endometrium, late proliferative phase, cycle day 13"			
/lab_host="DH10B (T1-resistant)"			
/clone_id="NICHDS_HS_Ut1"			
/note="Organ: uterus; Vector: pCMV-SPORT.1.cd4 (Reagen, Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV; Cloned unidirectionally from microquantity amounts of mRNA from normal endometrial tissue (late proliferative phase, cycle day 13). Average insert size 1.9 kb. Library constructed by Reagen (Invitrogen Corporation)."			
ORIGIN			
Query Match	33.5%	Score 622;	DB 6; Length 622;
Best Local Similarity	100.0%	Pred. No. 4.8e-284;	
Matches 622; Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1159	CGTGGACCCGGCTCTTCACTTCTTGAACACCCCAATGCGCAAGCTCTGAGGCTTCA	1218
Db	1	CGTGGACCCGGCTCTTCACTTCTTGAACACCCCAATGCGCAAGCTCTGAGGCTTCA	60
Qy	1219	GCAACGGCTCGGGGGACCCAGAGACGGAGCAAGCACTTTGTAGGTGCCCCCAAGGCCCA	1278

Db	Accession	Length	Score	E-value	Organism
Db	61	GCACGCGCTCCGGGGGACCCAGAGCGGGAGCACTTTGTGAGGTCCCAAGGCCCA	120		
Qy	1279	GGGCAAGAGTGTGGGTGTGGGCGCATGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGG	1338		
Db	121	GGGCAAGAGTGTGGGTGTGGGCGCATGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGG	180		
Qy	1339	ACCCGCCATCCCTGCGGGTGTCCCTCCGCGCACCTGTGGCCAGCCCGGCGCTCTCC	1398		
Db	181	ACCCGCCATCCCTGCGGGTGTCCCTCCGCGCACCTGTGGCCAGCCCGGCGCTCTCC	240		
Qy	1399	CTCCCTAGCCCTCTCGGGGCAAGAGACCAAGACCCAGCCCAAGAGAGCCAGCAGGG	1458		
Db	241	CTCCCTAGCCCTCTCGGGGCAAGAGACCAAGACCCAGCCCAAGAGAGCCAGCAGGG	300		
Qy	1459	CTGCGGGGGGCTGCGAGGCACCACTGAGCTCAGGTGGCCCTGTCTGTGGGGCGGAGCACT	1518		
Db	301	CTGCGGGGGGCTGCGAGGCACCACTGAGCTCAGGTGGCCCTGTCTGTGGGGCGGAGCACT	360		
Qy	1519	GCGGGAGCTGCGCCGCTTGTGTGTGTATGTAGAGCCAGGCGGGGAGCCGGTCCAGAGCA	1578		
Db	361	GCGGGAGCTGCGCCGCTTGTGTGTGTATGTAGAGCCAGGCGGGGAGCCGGTCCAGAGCA	420		
Qy	1579	TGAGCACACACACACACCATGTGACATGACACCATTAACACCATCTTACAGACATAGAG	1638		
Db	421	TGAGCACACACACACACCATGTGACATGACACCATTAACACCATCTTACAGACATAGAG	480		
Qy	1639	CCCTCTCCCGAGGGGCCACCCCTGCGCATGTAGAGAGCCCAACCCCGACACCAAGGAT	1698		
Db	481	CCCTCTCCCGAGGGGCCACCCCTGCGCATGTAGAGAGCCCAACCCCGACACCAAGGAT	540		
Qy	1699	TATTAATTCATTAATTAATTTGTATTAATGATTAATTTGTATTAATTAATTTGTACTC	1758		
Db	541	TATTAATTCATTAATTAATTTGTATTAATGATTAATTTGTATTAATTAATTTGTACTC	600		
Qy	1759	CACATAATTTAGCTAGCCTAC 1780			
Db	601	CACATAATTTAGCTAGCCTAC 622			

Db 582 TGTCCATGAGAGACAGCCGAGAGAGTGAACCTTCACTTGTATGACTTT 634

RESULT 14
BI767278 627 bp mRNA linear EST 25-SEP-2001
LOCUS 603057995F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207614 5',
DEFINITION
mRNA sequence.
ACCESSION BI767278 GI:15758869
VERSION BI767278.1 GI:15758869
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 627)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LLAM1520 row: m column: 23
High quality sequence stop: 580.

FEATURES
source
1..627
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5207614"
/lab_host="DH10B"
/clone_1lb="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: PCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

ORIGIN
Query Match 31.3%; Score 582; DB 4; Length 627;
Best Local Similarity 100.0%; Pred. No. 4.9e-265;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 765 GGTAAAGCTCACCCTGCGCCCGATGCGACCGACAGAGAGCGTCTTGTCAATCA 828
DB 1 GGTAAAGCTCACCCTGCGCCCGATGCGACCGACAGAGAGCGTCTTGTCAATCA 60
QY 829 GGCTGACCTGACAGAGGCGGCGGAGAGAGACCAAGCCACTGAGGAGCCTGGGAG 888
DB 61 GGCTGACCTGACAGAGGCGGCGGAGAGAGACCAAGCCACTGAGGAGCCTGGGAG 120
QY 889 CTGGGAGAGAGACAGCGAGCCCGCTCAGAGTTCAGAGGTGACGCGCTGGAGCAGTC 948
DB 121 CTGGGAGAGAGACAGCGAGCCCGCTCAGAGTTCAGAGGTGACGCGCTGGAGCAGTC 180
QY 949 TGGCTGCTACCACTTGGGTAGATGAGAACATCGAGAGAGAAACCACTTAATGATCT 1008
DB 181 TGGCTGCTACCACTTGGGTAGATGAGAACATCGAGAGAGAAACCACTTAATGATCT 240
QY 1009 CCGCGGAGTAGAAAATCAACGTCCTCAATTGGGCTGGCTCCCTTCCGTGGCCAGAA 1068

Db 241 CGCGGGAGTAGAAGAACTACAGTCCCAATTTGGGCTGCTCCCTTCCGTGGCCAGAA 300

QY 1069 GTGAAACTGCCCCCGGACCTTCAATCCACTGATCTGCTCCCATGAGCGGAAGC 1128
DB 301 GTGAAACTGCCCCCGGACCTTCAATCCACTGATCTGCTCCCATGAGCGGAAGC 360
QY 1129 CATCCATCCCAACACCGAAGGCCGAGCGTGAACCGGCTCTTCCACTTCTTGA 1188
DB 361 CATCCATCCCAACACCGAAGGCCGAGCGTGAACCGGCTCTTCCACTTCTTGA 420
QY 1189 CACCCCAATCGCAAGGTCTCAGAGTCCAGCAACGAGTCCGAGGACCCAGAGCGGAG 1248
DB 421 CACCCCAATCGCAAGGTCTCAGAGTCCAGCAACGAGTCCGAGGACCCAGAGCGGAG 480
QY 1249 CAGCACTTTGTGAGTCCCCCAAGGCCAGGCAAGAGTGTGGTGGCCACGTGC 1308
DB 481 CAGCACTTTGTGAGTCCCCCAAGGCCAGGCAAGAGTGTGGTGGCCACGTGC 540
QY 1309 CAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCAATCC 1350
DB 541 CAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCAATCC 582

RESULT 15
BG104777 1027 bp mRNA linear EST 30-JAN-2001
LOCUS 602311726F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4421335 5',
DEFINITION
mRNA sequence.
ACCESSION BG104777
VERSION BG104777.1 GI:12598623
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 1027)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LLAM10160 row: d column: 08
High quality sequence stop: 692.

FEATURES
source
1..1027
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4421335"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_84"
/note="Organ: adrenal gland; Vector: PCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 30.9%; Score 574; DB 4; Length 1027;
Best Local Similarity 99.8%; Pred. No. 3e-261;
Matches 624; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 287 AGCTTGCCGTGAGCGCTCTGCGCTCGAAGGCGATCGAGAGTGTGCGGAGACAG 346
DB 1 AGCTTGCCGTGAGCGCTCTGCGCTCGAAGGCGATCGAGAGTGTGCGGAGACAG 60

```
QY 347 CGCTGCCCGGGCGGTGCTCGGGAACCCCGACAAGCTGCGTGTGGCGGGCACATAGGCCGA 406
Db 61 CGCTGCCCGGGCGGTGCTCGGGAACCCCGACAAGCTGCGTGTGGCGGGCACATAGGCCGA 120
QY 407 AGCACCCGGAGAGCTCGTGGGCGACGTGTGTAGAGACACGCTCAGCGAGAGAGAGAGAC 466
Db 121 AGCACCCGGAGAGCTCGTGGGCGACGTGTGTAGAGACACGCTCAGCGAGAGAGAGAGAC 180
QY 467 GACTTTGGCTGGAAGTGGCCCTGCTCTGAGAAAGCTGACGGGCTGGGACCGAGAT 526
Db 181 GACTTTGGCTGGAAGTGGCCCTGCTCTGAGAAAGCTGACGGGCTGGGACCGAGAT 240
QY 527 GAGAAAGAAATGAGAGAGAGAGAACCTGCCAGGCTCCAAAGACAGCTGAAGTTT 586
Db 241 GAGAAAGAAATGAGAGAGAGAGAACCTGCCAGGCTCCAAAGACAGCTGAAGTTT 300
QY 587 GAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTGAACCTTCACC 646
Db 301 GAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTGAACCTTCACC 360
QY 647 CTGTATGACTTTTGAACAACAACGGCAGAGTCAACCCGAGAGACATCACACGCTTGCTGAC 706
Db 361 CTGTATGACTTTTGAACAACAACGGCAGAGTCAACCCGAGAGACATCACACGCTTGCTGAC 420
QY 707 ACCATCTATGAGGTGGTGGACTCCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTG 766
Db 421 ACCATCTATGAGGTGGTGGACTCCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTG 480
QY 767 CGGGTAAAGCTCACCGTGACCCCGATGGCAGCCAGACAAAGAGAGCGTCTTGTCAAT 826
Db 481 CGGGTAAAGCTCACCGTGACCCCGATGGCAGCCAGACAAAGAGAGCGTCTTGTCAAT 540
QY 827 CAGGCTGACCTTGCAGAGAGCGCAAGGCCCGAGCAGAGACCAAGCCACTGAGGACCTGCGG 886
Db 541 CAGGCTGACCTTGCAGAGAGCGCAAGGCCCGAGCAGAGACCAAGCCACTGAGGACCTGCGG 600
QY 887 AGCTGGGAGAAAGACAGCGAGGCC 911
Db 601 AGCTGGGAGAAAGACAGCGAGGCC 625
```

Search completed: December 30, 2004, 03:20:00
Job time : 8798 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 212 Seconds
(without alignments)
6232.814 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1 gaattcgccctctaactacg.....aaaaaaaaaagcgcgcc 1859

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1434.8	77.2	1438 4	US-09-506-066E-5 Sequence 5, Appl
2	1221.4	65.7	1731 4	US-09-506-066E-3 Sequence 3, Appl
3	252	13.6	1285 4	US-09-506-066E-7 Sequence 7, Appl
4	206.8	11.1	1307 4	US-09-506-066E-9 Sequence 9, Appl
5	57.6	3.1	7218 1	US-08-232-463-14 Sequence 14, Appl
6	53.8	2.9	3768 4	US-10-101-464A-949 Sequence 949, App
7	51.2	2.8	2790 4	US-09-254-594-5 Sequence 5, Appl
8	51.2	2.8	3432 4	US-09-254-594-4 Sequence 4, Appl
9	49.8	2.7	375 4	US-09-248-796A-7789 Sequence 7789, App
10	49.8	2.7	459 4	US-09-707-919A-13 Sequence 13, Appl
11	49.2	2.6	623 3	US-09-043-303-5 Sequence 5, Appl
12	48.4	2.6	5852 1	US-07-867-106-2 Sequence 2, Appl
13	48.4	2.6	4403765 3	US-09-103-840A-2 Sequence 2, Appl
14	48.2	2.6	516 4	US-09-083-268-1 Sequence 1, Appl
15	48.2	2.6	4481 3	US-09-041-886-18 Sequence 18, Appl
16	48.2	2.6	4481 4	US-09-648-281-1 Sequence 1, Appl
17	48.2	2.6	4481 4	US-09-707-919A-20 Sequence 20, Appl
18	48.2	2.6	4481 4	US-09-083-268-2 Sequence 2, Appl
19	47.6	2.6	3162 4	US-10-140-002-111 Sequence 111, App
20	47.4	2.5	4403765 3	US-09-103-840A-2 Sequence 2, Appl
21	47.4	2.5	4411529 3	US-09-103-840A-1 Sequence 1, Appl
22	47.2	2.5	1859 3	US-09-359-913-23 Sequence 23, Appl
23	47.2	2.5	1859 3	US-09-298-731-23 Sequence 23, Appl
24	47.2	2.5	1859 3	US-09-350-614-23 Sequence 23, Appl
25	47.2	2.5	1955 3	US-09-399-913-19 Sequence 19, Appl
26	47.2	2.5	1955 3	US-09-298-731-19 Sequence 19, Appl
27	47.2	2.5	1955 4	US-09-350-614-19 Sequence 19, Appl

28	47.2	2.5	2009 3	US-09-399-913-13 Sequence 13, Appl
29	47.2	2.5	2009 3	US-09-238-731-13 Sequence 13, Appl
30	47.2	2.5	2009 4	US-09-350-614-13 Sequence 13, Appl
31	46.6	2.5	430 4	US-09-621-976-16656 Sequence 16656, A
32	46.4	2.5	7055 4	US-09-976-594-941 Sequence 941, App
33	46	2.5	435 4	US-09-389-681-285 Sequence 285, App
34	46	2.5	435 4	US-09-620-405B-285 Sequence 285, App
35	46	2.5	435 4	US-09-339-338-285 Sequence 285, App
36	46	2.5	435 4	US-09-433-826B-285 Sequence 285, App
37	46	2.5	435 4	US-09-604-287A-285 Sequence 285, App
38	46	2.5	435 4	US-09-834-759-285 Sequence 285, App
39	46	2.5	435 4	US-09-590-751A-285 Sequence 285, App
40	46	2.5	1149 3	US-09-422-487-6 Sequence 6, Appl
41	46	2.5	1149 3	US-09-378-088A-128 Sequence 128, App
42	46	2.5	1149 4	US-09-643-596B-128 Sequence 128, App
43	46	2.5	1196 4	US-09-270-767-14804 Sequence 14804, A
44	45.6	2.5	152331 3	US-09-158-155-16 Sequence 16, Appl
45	45.4	2.4	528 4	US-09-232-785-242 Sequence 242, Appl

ALIGNMENTS

```
RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
APPLICANT: Zeng, Wenlin
TITLE OR INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/120,646
PRIOR FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8)...(1418)
OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-5

Query Match 77.2%; Score 1434.8; DB 4; Length 1438;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

217 CCCAGATGGGAACTTCACTCCAGCCGCGCGCTGCAAGGCGAGGAGAGCC 276
1 CCCAGATGGGAACTTCACTCCAGCCGCGCGCTGCAAGGCGAGGAGAGCC 60
277 GGAAGTGAAGCTTCCCGTGAAGCGCTGCGAGTGGAGTGAAGGAGTGAAT 336
61 GGAAGTGAAGCTTCCCGTGAAGCGCTGCGAGTGGAGTGAAGGAGTGAAT 120
337 CGGAGAGCAGCGCTGCGCGGCGGTGTCTCGGAGACCCGACAGCTGGCGGCGAC 396
121 CGGAGAGCAGCGCTGCGCGGCGGTGTCTCGGAGACCCGACAGCTGGCGGCGAC 180
397 CATGAGCGGAAGCAGCGGAGCGTGGGCGAGCTGTGAGAGACAGCTCAGCAGGA 456
181 CATGAGCGGAAGCAGCGGAGCGTGGGCGAGCTGTGAGAGACAGCTCAGCAGGA 240
457 AGAGAGAGAGCAGCTTTCGAGTGAAGTGGCTTCTTGAAGAGCTGAAGGCTGGG 516
241 AGAGAGAGAGCAGCTTTCGAGTGAAGTGGCTTCTTGAAGAGCTGAAGGCTGGG 300
```

QY 517 CAGCGGATGAGAGAGAGATGAGAGAGGCGAACCTGCCAGGCTCCAGAGCA 576
DB 301 CAGGGAGATGAGAGAGAGATGAGAGAGGCGAACCTGCCAGGCTCCAGAGCA 360
QY 577 GCTGAATTTGAGAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG 636
DB 361 GCTGAATTTGAGAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG 420
QY 637 GACCTTCACTCTGTATGACTTTTGACCAACACGGCAGAGTCACTCCGAGAGACATCACCG 696
DB 421 GACCTTCACTCTGTATGACTTTTGACCAACACGGCAGAGTCACTCCGAGAGACATCACCG 480
QY 697 CTTGCTGACACCATCTATGAGAGGTGAGACTCCTGTCAACCATCCCCAATCATCAG 756
DB 481 CTTGCTGACACCATCTATGAGAGGTGAGACTCCTGTCAACCATCCCCAATCATCAG 540
QY 757 CAAGATCTGGGGTAAAGCTCACCGTGGCCCCCGATGCGACGCGACGACAGAGAGAGCT 816
DB 541 CAAGATCTGGGGTAAAGCTCACCGTGGCCCCCGATGCGACGCGACGACGAGAGAGAGCT 600
QY 817 CTTGTCAATCAAGCTGACCTGACAGAGCGCAAGGCCCCGAGACAGACCAAGCCACTGA 876
DB 601 CTTGTCAATCAAGCTGACCTGACAGAGCGCAAGGCCCCGAGACAGACCAAGCCACTGA 660
QY 877 GGAACCTGGAGCTGGAGAGAGAGAGAGAGCGAGCCCGCTCAGGTTCCAGGGTGA CAGCGG 936
DB 661 GGAACCTGGAGCTGGAGAGAGAGAGAGAGCGAGCCCGCTCAGGTTCCAGGGTGA CAGCGG 720
QY 937 CTTGAGACAGTGTGGCTGCTACACCATTTGGTATGAGAGACATCGAGAGAGAGAGCA 996
DB 721 CTTGAGACAGTGTGGCTGCTACACCATTTGGTATGAGAGACATCGAGAGAGAGAGCA 780
QY 997 CTATTGATCTCGCCGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1056
DB 781 CTATTGATCTCGCCGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 1057 CGTGGCCAGAGAGCAG 1116
DB 841 CGTGGCCAGAGAGCAG 900
QY 1117 TGAACCGGAGAGCAG 1176
DB 901 TGAACCGGAGAGCAG 960
QY 1177 CCACTTCTTGAACAG 1236
DB 961 CCACTTCTTGAACAG 1020
QY 1237 CAG 1296
DB 1021 TCAAG 1080
QY 1297 GGGGCACTGGGCAAG 1356
DB 1081 GGGGCACTGGGCAAG 1140
QY 1357 GTCCCCCTCGGCACTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416
DB 1141 GTCCCCCTCGGCACTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1417 GCAAG 1476
DB 1201 GCAAG 1280
QY 1477 ACCACTGGGCTCAGTGGGCTGTCTGGGGCGGAGAGACCTGGGAGAGCTGCCGCTT 1536
DB 1261 ACCACTGGGCTCAGTGGGCTGTCTGGGGCGGAGAGACCTGGGAGAGCTGCCGCTT 1320
QY 1537 GGTGGTGTATGAG 1596
DB 1321 GGTGGTGTATGAG 1380
QY 1597 TGAACATCAACCATTTACAGCACTTCTACAGACATAGAGGCCCTCCCAAGGAGCC 1654

DB 1381 TGAACATCAACCATTTACAGCACTTCTACAGACATAGAGGCCCTCCCAAGGAGCC 1438

RESULT 2
US-09-506-066E-3
Sequence 3, Application US/09506066E
Patent No. 6610323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
APPLICANT: Zeng, Wenlin
TITLE OF INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/120,646
PRIOR FILING DATE: 1999-02-17
NUMBER OF SEQ. ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1731
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (140)...(1553)
OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-3

Query Match 85.7%; Score 1221.4; DB 4; Length 1731;
Best Local Similarity 83.9%; Pred. No. 3,3e-267;
Matches 1460; Conservative 0; Mismatches 256; Indels 25; Gaps 6;

QY 82 CGGCGGAGAGAGCGGCTCCGCGCGCGCTCGGAGCTCCGCTCGGAGCTGGGAGCTCTTCG 141
DB 6 CGGCGGAGAGAGCGGCTCCGCGCGCGCTCGGAGCTCCGCTCGGAGCTGGGAGCTCTTCG 63
QY 142 GAG 201
DB 64 GAG 121
QY 202 TCCGCGCGCGGAG 261
DB 122 CCGCGCGCGC-----GCCCGAGATGGGAGAACTTCACTCGAAGCGGCGCGCTGTCAA 177
QY 262 GCGGAG 321
DB 178 GCGGAG 237
QY 322 CATGAGAGAGATGATCGGAG 381
DB 238 CATGAGAGAGATGATCGGAG 297
QY 382 GCGGTTGGCGGAG 441
DB 298 GAGATTGGCAGGAG 357
QY 442 CAGGCTCAGGAG 501
DB 358 GGCCTCTCGTGAAG 417
QY 502 GACTGAGCGGCTGGGAG 561
DB 418 GATCGAAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 477
QY 562 AGGCTCAAG 621
DB 478 GGCCTCAAG 537
QY 622 CAGCGGAG 681
DB 538 CAGCGGAG 597

QY 682 AGAGACATCAACGCTTGTGTCACACCATCTATGAGTGTGAGCTTCTCTCAACCA 741
 DB 598 TGAAGCATTTACACGCTTGTGTCATACATCTAAGAGGTGATCTCTGTGAACCA 657
 QY 742 CTCCTCAACATCCAGCAAGATGTGGGTAAAGCTACCGTGGCCCCCGATGGCAGCCA 801
 DB 658 TTCTCCCATCATCANAGCAAGACATGCGGTGAACTCAACGTGCTCTGACGGAGACCA 717
 QY 802 GAGCAAGAGAGCGCTCTGTGTCATCAGGTGACCTGACAGAGCCAGCAAGCCCCGAGACA 861
 DB 718 GATGTAAAGAGAGGTCTTTTCAACATACCGATCTGACAGACCAAGGCCCGAGACAGA 777
 QY 862 GACCAAGCCCACTGAGGACCTGCGGAGCTGGAGAGAAAGACGAGCCCCGCTCAGGTT 921
 DB 778 CACCAAAACCCGCTGAGAGAGCTGCGTGTGGAGAAAGACGAGCCCCCAGCTCAGGTT 837
 QY 922 CAGAGGTGACAGCGCGCTGAGAGAGTGTGGTGTACACCATTTGGCTGATGAGAACAT 981
 DB 838 CAGAGGTGACAGCGCACTGGAGAGCCAGACTGTCTACACCATTTGGCTGATGAGAACAT 897
 QY 982 CGAGAGAGAAACCACTATTAGATCTCGCGGATAGAAAATPACAGCTCCCAATTTGG 1041
 DB 898 TGAAGAGAGAAACCACTATTAGATCTCGCGGATAGAAAATPACAGCTCCCAATTTGG 957
 QY 1042 GCGTGTGCTCTCTGCTGCGTGGCCAGAAAGTCAAACTGCCCCCGCAGCTTCCAACTCCAC 1101
 DB 958 ACCGGAGTCCCTTCCGCTGCGTGGCCAGAAAGTCAAACTGCCCCCGCAGCTTCCAACTCCAC 1017
 QY 1102 TCGATCTCGCTCCCATGAGCGGAGGCAATCCACATCCACAGCAAGGCCCAAGGCGT 1161
 DB 1018 TCGCTCTCGCTCCCATGAGCGGAGGCAATCCACATCCACAGCAAGGCCCAAGGCGT 1077
 QY 1162 GGAAGCGGCGCTCTTCCATCTCTTGAACACCCCAATGCGCAAGGCTCTCAGAGCTCCAGCA 1221
 DB 1078 GGAAGCGGCGCTCTTCCATCTCTTGAACACCCCAATGCGCAAGGCTCTCAGAGCTCCAGCA 1137
 QY 1222 ACGGCTCTCGGGGCAACCAAGAGCGGAGCAAGCACTTTGTGAGTCTCCCAAGGCCCAAGG 1281
 DB 1138 ACGGCTCTCGGGGCAACCAAGAGCGGAGCAAGCACTTTGTGAGTCTCCCAAGGCCCAAGG 1197
 QY 1282 CAGAGGTGTGGGTGTGGGCAACGAGCGGAGGAGCAAGCAAGGCCCTCTGGGAGC 1341
 DB 1198 CAGAGGTGTGGGTGTGGGCAACGAGCGGAGGAGCAAGCAAGGCCCTCTGGGAGC 1257
 QY 1342 CGCCATCTCTGCGGTGTGCTCTGCGGCACTGTGCTGCAAGCGGCGCTCTCTCCCTC 1401
 DB 1258 CAGCAACCATATCTGTCTCTCTCTGCGCACTGTGCTGCAAGCGGCGCTCTCTCCCTC 1317
 QY 1402 CTTAGCCCCCTTGGGCAACAAGACCAAGACCGGAGGAGGAGGAGGAGGAGGAGGAGG 1461
 DB 1318 CTTAGCCCCCTTGGGCAACAAGACCAAGACCGGAGGAGGAGGAGGAGGAGGAGGAGG 1377
 QY 1462 CCGGGGCTGTGAGGCAACATGCGCTCAGGTG---CCGTCTCTGGGGGGGAGGAGCCT 1518
 DB 1378 CCGGGGCTGTGAGGCAACATGCGCTCAGGTG---CCGTCTCTGGGGGGGAGGAGCCT 1437
 QY 1519 GCGGAGCTGCGCGCTTGT 1578
 DB 1438 GAGGAGCTGCGCGCTTGT 1497
 QY 1579 TGAAGCAACCAACATGAGAACATCAACCACTTACCACTTGTACAGACATGAG 1638
 DB 1498 CGAAACATCATACCAACCAACATCAACCACTTATACCACTTGTACAGCCTTGTAGC 1557
 QY 1639 CCGCTCCCAAGGCGCCACCTGTCATATGAAAGAACCCCGGACACCAAGGAGCAT 1698
 DB 1558 CC-----CAGCAGGTGTCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1603
 QY 1699 TATTATTCTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1758
 DB 1604 TATTATTCTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1663

QY 1759 CACTAATATTAGTACGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 1818
 DB 1664 CACTAATATTAGTACGCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 1723
 QY 1819 T 1819
 DB 1724 T 1724
 RESULT 3
 US-09-506-066E-7
 ; Sequence 7, Application US/09506066E
 ; Patent No. 6630323
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, Matthew
 ; APPLICANT: Wharton, Keith
 ; APPLICANT: Zeng, Wenlin
 ; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
 ; FILE REFERENCE: STAN-121
 ; CURRENT APPLICATION NUMBER: US/09/506,066E
 ; CURRENT FILING DATE: 2000-02-17
 ; PRIOR APPLICATION NUMBER: 60/120,646
 ; FILING DATE: 1999-02-17
 ; NUMBER OF SEQ. ID NOS: 15
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1285
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (2)...(1282)
 ; OTHER INFORMATION: Nkd2 coding sequence
 US-09-506-066E-7
 Query Match 13.6%; Score 252; DB 4; Length 1285;
 Best Local Similarity 56.3%; Pred. No. 1,3e-47;
 Matches 631; Conservative 0; Mismatches 450; Indels 39; Gaps 7;
 QY 516 GCAAGCGAGATGAGAAAGATGAGAGAGTGAAGCAAGCCCTGCCAGCTCCAGAAAGC 575
 DB 180 GCAAGATGATGGGAG 239
 QY 576 AGCTGAAGTTTGAAGAGCTTCAAGTGCAGCTGTTCATGTAGAGAGAGAGAGAGAGAG 635
 DB 240 ACTTGAACATTGACGACATGACAGTGTGACCTCTCAGTGAAGAGAGAGAGAGAGAG 299
 QY 636 GGAAGCTTCAACCTGTATGATTTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 695
 DB 300 GGAAGCTTCAACCTGTATGATTTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
 QY 696 GCTTGTGACACCAATCTATGAGGTGTGAGTCTCTGTCAACCACTCTCCCAATCA 755
 DB 360 GCTTGTGACACCAATCTATGAGGTGTGAGTCTCTGTCAACCACTCTCCCGGAGCA 419
 QY 756 GCAAGATGCTGCGGTAAAGCTCAACGTGAGCCCCCGATGCGAGCGAGAGAGAGAGAG 815
 DB 420 GCAAGATGCTGCGGTAAAGCTCAACGTGAGCCCCCGATGCGAGCGAGAGAGAGAGAG 476
 QY 816 TCTTGTGATCAAGGCTGACCTGACAGAGCGCAAGGCCCGAGAGAGAGAGAGAGAGAG 875
 DB 477 GTCTCTCACTGAGCAAG 536
 QY 876 AGGACCTGCGAGAGCTGGAG 935
 DB 537 ATGAGGCCGAG 596
 QY 936 GCTTGAAGAGTGTGCTGTACCACTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 995
 DB 597 CCAAGCCCTGCTGTGCGAGTGCCTTACGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAG 656
 QY 996 ACTATTAGATCTGCGCGGATGAGAAACTACAGTCCCAATTTGGGCTGCTCCCTT 1055

Dh 657 ACTACCTAGACCTTGTGGATCGAGAACTACATCATTAAGTTTGTGCTGGGTCAAC 716
Qy 1056 CCGTGGCCCAAGAACTCAAGAACTGCCCCCGGCACCTC---CAATCCCATCTGATCTGCT 1112
Db 717 CTGAGCGAGCGCAGCAAGAACTATGAGCAGGCGCACACATTCACAGAGGTCCCGAT 776
Qy 1113 CCCATGAGCGGGAAGCCATCCACATCCCAACCGAAAGCCCAAGGGGTGGACCCGGGCT 1112
Db 777 CACAGAGTGGATGCGCCACGCTATACACACCGCAGGTCTCAAGTCTGG----- 827
Qy 1173 CTTTCCACTTCTTGACACCCCAATCGCAAGGTCTGAGAGCTTCAGACAAAGGCTCGGG 1232
Db 828 --CTGACATGTCATATACAGTATAGCTTGACCCCGGCGCTGGGTGACAGCCCG 885
Qy 1233 GCACCCAGAGCGGAGCAAGCATTTTGTAGAGTCCCCCAAGGCCCAAGGCAAGAGTGG 1292
Db 886 GATCAAGGGGCGAGAG--AAGCAGTTCTCAGGTCTCTTAAGGGTCAAGAAACCTCT-- 942
Qy 1293 GTGTGGGCGACGTGGCCAGAGGGGCAAGAAACAGACCCCTCTGGGACCCGCGATCCCTG 1352
Db 943 ---TGGACACACAGGCGAGTGGCAAGCCAGGAAAGCTTCAAGTATTTGCTGAGGCGCT 999
Qy 1353 CGGTGTCCCTCCGCCACCTGCTGCGAGCCCGGCTCTCTCCCTCCCTAGCCCCC 1412
Db 1000 GCCATGCCCCAGAGTGTCTAGATGGCACACCTTCTCAGCCCGCACAGACGCTCC 1059
Qy 1413 TCGGGCAAGAGACCAAGACCGAGCCAGAGAGCCAGAGGCTGCGGGGCTGC 1472
Db 1060 ACCCGACCCCTATGTGCA-----CAAGCGGTACCGGCAAGAAAGCCAGAAAGGCC 1109
Qy 1473 AGGCAACCTGCGCTCAGGTGGCC-----CTGTCTGAGGGCGGAGACCTCGGGAGC 1526
Db 1110 ACTTACCACTTAAGGGGCAATGACCACTTACATGATGAGACATGAAGTACTAGGAGCC 1169
Qy 1527 TGCCCGCTTGTGTGTATGAGCAAGCCGCGGCAAGCCGCTCCAGAGCATGAGCAAC 1586
Db 1170 TGCTCCCATGCTGGGGCTGAGGGCTATGTGATGCTGTGGTCTCAGAGGATGAACACC 1229
Qy 1587 ACCACCACTGAATCACCACCATTAACCACTTCTA 1626
Db 1230 ACCATCACATGAGACACCAACCATCAACCAACCA 1269

RESULT 4
US-09-506-066E-9
; Sequence 9, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-506-066E-9

Query Match 11.1%; Score 206.8; DB 4; Length 1307;
Best Local Similarity 55.1%; Pred. No. 2.2e-37;
Matches 553; Conservative 0; Mismatches 412; Indels 39; Gaps 6;

Qy 670 CAAGGTCAACCCGAGAGACATCACCTTGTGTCACACCATCTATAGGTGTGAGCTC 729
Db 51 CGAGGTCAACGAGGAGACATGTCCAGGCTCATGACACCACTATATGAGGTGTGATGC 110

Qy 730 CTCGTCAACACGATCCCAACATCCAGCAAGATGTGCGGGTAAAGCTACCGTGGCCC 789
Db 111 CTGGTAAACCACTTCTCGGGCAGCAGCAAGACCTTCCGTGTGAAGTTAACGTCAGCC 170
Qy 790 CGATGCGACGACAGCAAGAGAGCGCTTGTCAATCAGGCTGACCTGACAGCGGAG 849
Db 171 TGAGCCCTCCAGCAAGAGAGAGAGGCTCTCTGTGGCCAGGACCGGAGGCCACCG 230
Qy 850 GCCCGAGCAGAACCAAGCCCATGAGGACTGCGGAGCTGGGAGAGAGAGAGCAGC 909
Db 231 TTGCAGATGAGAGGTGAACCTGACAGAGCCAAAGGTGGCTGACAGAGTTGTCTGC 290
Qy 910 CCGGCTCAGGTTCCAGGGGTGACAGCGCTGAGAGCTGTGCTGTACCAACCATTCGT 969
Db 291 ACAAGTGAAGGCCCAAGTACTGACCTCCAGCCCTGTGCGAGCGGGGCTTACTGCT 350
Qy 970 AGATGAAACATGAGAGAGAGAAACCACTTATAGATCTGCGGGATAGAAACTACAC 1029
Db 351 GGAAGAGAACAGGAGGAGAAACCACTACTGACCTGCGGGATGAGAACTACAC 410
Qy 1030 GTCCCAATTTGGGCTGTGCTCCCTTCCGTGGCCCAAGATCAGAACTGCCCCCGCAC 1089
Db 411 GTCCAGATTTGGGCTGTGGTCCCTCTCTGTGCAAGCAAGAGAGCCAGGGCAAGGC 470
Qy 1090 CTCCAATCCC---ACTGATCTGCTCCATGAGCCGGAAGCCATCCATCCACACCG 1146
Db 471 CTGCACTTCAAGGCTCGGTGTCGCTCCGCTCCAGAGCCAGATACATGCTTACACACG 530
Qy 1147 AAGCCCAAGGGGTGAGCCCGGCTCTTTCATCTTGTGACACCCCAATGCGAAGT 1206
Db 531 CAGGTCAAGGTCTGGTGAACACGTCGAGCCTCGAGACCTGCG--TGCCGGGC 587
Qy 1207 CTGAGACTCAAGCAAGCTCCGGGGCACCCAGAGCGGAGAGCAAGCATTTGTAGTGC 1266
Db 588 CTGGAACACGACCCCGGCGAAGGGCGGAG-----AAGAGTTCTCAAGTGC 638
Qy 1267 CCCCAAGGCCCAGGGCAAGAGTGGGTGTGGGCCACAGTGGCCAGAGGGGCAAGAAACA 1326
Db 639 CCCAAGGGCTCGGGAAAGCCGCTGGGGTGGCCAGCAGACAGTCCGGGAAAGCTT 698
Qy 1327 GCCCCTCTGGAACCCGCAATCCCTGCGGTGCTCCCTCCGCACTGCTGCGAGCCC 1386
Db 699 CAGCTACTGCTGCGGCGCTGC-----TGCGGCCAGGGCCCTCAGAGCGGCACCA 752
Qy 1387 GGCCTCTCCCTCTCTTAAGCCCCCTCGGGCAAGAAACCAAGACCGAGCAAGGA 1446
Db 753 CTTCGCGAAGCCCAACCGCACCTTACGCGCAACAGCGGTACCGCAAAAGGCAAGGA 812
Qy 1447 GAGCAGCAAGGCTGCGGGGCTGACAGCACACTGAGGTGAGGCTGTCTGGG 1506
Db 813 GGGCACTGCGCACTCAAGGCCCAACAGCTCAGCTTGCACAG-----TGGA 860
Qy 1507 GCGGAGCACTGCGGAGGTGCTCCGCTTGTGTGTATGAGAGCCAGGCCGAGGCC 1566
Db 861 GCAAGAGTGTGCGGGACTGCGCGCCAGCGAGAGAGGAGGTCTACCGGTGCGAGT 920
Qy 1567 GGTCCAGAGCATGAGACCAACCAACCACTGAATCATCACCACATTT-----ACCA 1620
Db 921 GATTCAGCGGCAAGAGACCAACCAACCAAGAGCAACACCAACCAACCAACCA 980
Qy 1621 CTTCTACCAAGATAGAGCCCTCCCGAGGGGCCCACTGCGCA 1664
Db 981 CTTCCACCGTCTTAAGGCCACTGCGCAAGCAACTGCTCCCA 1024

RESULT 5
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.


```

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fl8
; US-08-232-463-14

Query Match      3.1%; Score 57.6; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. No. 0.003;
Matches 21; Conservative 225; Mismatches 164; Indels 0; Gaps 0;

QY 308 TGGGCTCGAAGCGCATCGAGTGTGCGAGAGACGCGCTCCGCGGTGTCTCG 367
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1441 TTGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

QY 368 GGAACCCGACAGTCGCGTGGCGCGGACACATAGCGGAGACCCGAGAGCTCGTGGC 427
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

QY 428 GAGGTGTGAGACAGCTCAGCGAGAGAGAGAGACACTTCGCTGGAAGTGGCC 487
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

QY 488 CTCGCTCTGAGAGACTGAGCGGTGCGAGCGAGATGAGAGAGATGAGAGAGTG 547
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

QY 548 AGCGAACCTCGCCAGGCTCCAGAGACAGCTGAAGTGTGAGAGCTCCAGTGGAGCTG 607
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

QY 608 TCATGAGAGAGAGACAGCCGCGAGAGTGAACCTTCACTGTATGACTTGAACAAGAC 667
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082

QY 668 GGAAGGTCAACCGAGAGACATCAACAGCTTGTGACACCATCTATGA 717
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1081 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1032
```

```

RESULT 6
US-10-101-464A-949
; Sequence 949, Application US/10101464A
; Patent No. 6768041
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 949
; LENGTH: 3768
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-101-464A-949
```

```

Query Match      2.9%; Score 53.8; DB 4; Length 3768;
Best Local Similarity 89.2%; Pred. No. 0.016;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```

QY 13 CTATAGCACTCACTATGAGGCAAGCAGTGTAAACGACGAGTACGCGGAGTGGG 72
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 CTATAGCACTCACTATGAGGCAAGCAGTGTAAACGACGAGTACGCGGAGTGGG 60
```

```

QY 73 CCGCG 77
    |
DB 61 ATGAG 65
```

```

RESULT 7
US-09-254-594-5
; Sequence 5, Application US/09254594
; Patent No. 6566094
; GENERAL INFORMATION:
; APPLICANT: KIMURA, Toru
; APPLICANT: KIKUCHI, Kaoru
; TITLE OF INVENTION: NOVEL SEMAPHORIN GENE: SEMAPHORIN Y
; FILE REFERENCE: 0020-4527P
; CURRENT APPLICATION NUMBER: US/09/254,594
; CURRENT FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: Tissue Type: Child Brain
; NAME/KEY: CDS
; LOCATION: (1)..(2790)
; OTHER INFORMATION: Identification Method: E
; NAME/KEY: misc.feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: Identification Method: P for resulting peptide
US-09-254-594-5
```

```

Query Match      2.8%; Score 51.2; DB 4; Length 2790;
```



```

Query Match      2.6%; Score 48.4; DB 1; Length 5852;
Best Local Similarity 57.1%; Pred. No. 0.33;
Matches 88; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1697 ATTATTTTCATTAAATTAATGTTATATATGATGATTAATGTTAATTAATTAATTTGTTAC 1756
Db 5753 ATTGGATTTTTTTTAAATATGTTATTTGTTGTTGTTTTTTAACTTAATTTTCATTTTT 5694

QY 1757 TCCACTAATATTAGCTAGCCTACATGTAGAGATCTATGGAAAACAGAACTAACTTT 1816
Db 5693 TATTTAATAAATTTAATTAATTAATTTTAATTAATAAAAAAAAAAAAAAAAAATTT 5634

QY 1817 TATTTATATGTTAAAAAAAAAAAAAAAAAAAAA 1850
Db 5633 AAATTAATAAAAAAAAAAAAAAAAAAAAAA 5600

RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      2.6%; Score 48.4; DB 3; Length 4403765;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 121; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 70 GGGCGCGCGCAAGCGCGGCAAGAGCGCTCCCGGCGCGCTCGGGCTCGGCTGCTG 129
Db 1637226 GGGCGCGCGTCCGCTGTACCGCGGTGCGGTGGGGTTCGCGGGTTCGGCGGGCGG 1637167

QY 130 GGGGCTGCTTGGAGAGAGAGAGAGAGGCGCCAGGCGCGCGGCGCGGCGCATG 189
Db 1637166 CGGCTGATACGCTGCTGGGCGGCGCAAGCGCGGCGGGGCGCGGAGACGGGCGGC 1637107

QY 190 GCTTAGGAGCGCTCCCGGCGCGCGAGCCCGAGATGGGAAACTTCACTTCAAGCCGCG 249
Db 1637106 TGTGTGCGTGGCGGATGACCGCGGAGCGGCGCAACGAGCGCTTCAACGAGG 1637047

QY 250 CGCGCTGTGCAAGCGGAGAGAGAGCGCGGAAAGGTGAACAGCTTCCGCGTGAAGCGTGGCTG 309
Db 1637046 TGC CGCGCGGTGACGCGCGGCAACGCGCGCAACCCCGGACCGGCGGAGCGACCGG 1636987

QY 310 GG 311
Db 1636986 CG 1636985

RESULT 14
US-09-083-268-1
; Sequence 1, Application US/09083268
; Patent No. 667353
; GENERAL INFORMATION:

```

```

APPLICANT: Pulet, Stefan M
TITLE OF INVENTION: NUCLEIC ACID ENCODING SPINOCEREBELLAR
TITLE OF INVENTION: ATAXIA-2 AND PRODUCTS RELATED THERETO
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Muehling, Raasch & Gebhardt, P.A.
STREET: 119 No. 6673535th Fourth Street
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,268
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/727,084
FILING DATE: 08-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: McCormack, Myra H
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 232,00010101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/305-1220
TELEFAX: 612/305-1228
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
US-09-083-268-1

Query Match 2.6%; Score 48.2; DB 4; Length 516;
Best Local Similarity 51.6%; Pred. No. 0.12;
Matches 145; Conservative 0; Mismatches 123; Indels 13; Gaps 1;

QY 1336 GGAACCCGCCATCCCTGGGGTGTCCCTCCCGCCACCTGCTGCCAGACCCTGGCTCTT 1395
DB 66 GGGCTCTCCCGCCCTTCTGTCGTGTCCTTCTCCCTCGCAACCCGGGGCTCTCG 125
QY 1396 CCCCTCCCTGAGCCCCCTCGGGGCAAGAAGCAACAGACCCGACCAAGAGACGACGA 1455
DB 126 GCCCGCGCAACCCCGCGC-----CTCCCCGTGGCCCGCGTGGTCTCCGCC 172
QY 1456 GGGCTGCGGGGGCTGAGGAGGACCACTGGCTTCAGTGGCCCTGCTCTGGGGCGGAGCA 1515
DB 173 GCGTTCGGGCGCTCTCTTGGGGCGCCCGGCTCCCGGCTGCCCGCGGGGTGCGAGCC 232
QY 1516 CCTGCGGAGCTGCCGCTTGTGTGTATGAGAGCCGAGCGGAGACCGCTGTCAAG 1575
DB 233 GGTGTATGGGCCCTCACTCATGTGCTGAAGCCCAAGACGAGAGGAGGACGACGACA 292
QY 1576 ACATGAGCACACACACACCATTAACATCAACCAACATTAC 1616
DB 293 ACAGCAGACGACGACGACGACGACGACGACGACGACGACG 333

RESULT 15
US-09-041-886-18
; Sequence 18, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
;

```

NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..4099
US-09-041-886-18

Query Match 2.6%; Score 48.2; DB 3; Length 4481;

Best Local Similarity 51.6%; Pred. No. 0.32;

Matches 145; Conservative 0; Mismatches 123; Indels 13; Gaps 1;

QY	1336	GGGAGCCGCGCATCTGCGGCTGTCCTCCGCGCCACCTGAGTGCAGCCCGGCTCTCT	1395
DB	457	GGGCTTCCTCCGCGCTTCGCTGCTGCTCTCTCCCTCGCAGCCCGGCGCTCTCG	516
QY	1396	CCCTCTCTTAGCCCTCCCTCGGCAACAAGACCAAGCAGCCGAGCCAAAGAGAGCCAGCA	1455
DB	517	GCCGCGCAACCGCGC-----CTCCCGCTCGGCGCCCGTCTGCTCCCGCC	563
QY	1456	GGGCTGCGGCGGCTGAGGACACCTGAGCTCAGTGGCTCTGCTCTGGGCGGAGCA	1515
DB	564	GCGTTCGCGGCTCTCTTGGCGGCGCCGCTCCCGCTGTCCTCCGCGGCTGCGAGCC	623
QY	1516	CTGCGGAGACTCGCGCTTGGTGTATGAGAGCCAGCCGCGAGCCGATCCAGAG	1575
DB	624	GCTGTATGGGCTCTCAGCATGTGCTGAAGCCCGCAGCAGCAGCAGCAGCAGCA	683
QY	1576	ACATGAGACACACACCATGAAATCAACCATTTACC	1616
DB	684	ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	724

Search completed: December 29, 2004, 21:00:03
Job time: 224 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 12708 Seconds
(without alignments)
6917.812 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gaattcgccctctaactg.....aaaaaaaaaagcgcgcg 1859

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank1:*
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1786	96.1	1786	9	AY061883 Homo sapi
2	1763	94.8	2604	9	BC051288 Homo sapi
3	1610	86.6	1788	9	AB062886 Homo sapi
4	1438	77.4	1438	9	AF358135 Homo sapi
5	1434.8	77.2	1438	6	AR405917 Sequence
6	1259.6	67.8	1765	10	BC034838 Mus muscu
7	1221.4	65.7	1731	6	AR405916 Sequence
8	1194.8	64.3	1671	10	AF358134 Mus muscu
9	1078.4	58.0	1416	10	AF343352 Mus muscu
10	1066.6	57.4	1401	6	AX172351 Sequence
11	849.2	45.7	1053	6	C0721681 Sequence
12	783	42.1	163319	9	HSA303140 Homo sapi
13	783	42.1	169732	9	AC007608 Homo sapi
14	783	42.1	170791	9	AC007334 Homo sapi
15	775	41.7	178790	2	AC145238 Homo sapi
16	754.2	40.6	194869	2	AC145175 Pan trogl
17	754.2	40.6	194869	2	AC145003 Papio anu
18	754.2	40.6	198189	2	AC145043 Papio anu
19	637.8	34.3	139145	2	AC146326 Felis cat

20	634.8	34.1	199277	2	AC145445 Canis fam
21	602	32.4	181401	2	AC146663 Sus scrof
22	592	31.8	208061	2	AC145254 Bos tauru
23	560.4	30.1	232509	2	AC134093 Rattus no
24	558.8	30.1	240965	2	AC126054 Mus muscu
25	272.4	14.7	1682	9	AF358137 Homo sapi
26	272.4	14.7	1798	9	AB062887 Homo sapi
27	255	13.7	1967	10	BC019952 Mus muscu
28	254.4	13.7	1947	10	AF358136 Mus muscu
29	252	13.6	1285	6	AR405918 Sequence
30	245.4	13.2	1885	6	AX780854 Sequence
31	232.8	12.5	1309	6	BC004940 Homo sapi
32	207	11.1	1958	9	BC012176 Homo sapi
33	206.8	11.1	1307	6	AR405919 Sequence
34	195.8	10.5	212531	2	AC145178 Gallus ga
35	169.2	9.1	1191	6	C0715890 Sequence
36	137	7.4	1950	6	AX557538 Sequence
37	136.6	7.3	2340	6	C0741468 Sequence
38	101.8	5.5	237332	2	AC131846 Rattus no
39	85	4.6	125020	9	AF429315 Homo sapi
40	75	4.0	99395	2	AC010446 Homo sapi
41	75	4.0	187289	9	AC116351 Homo sapi
42	73.4	3.9	181105	5	BX664719 Zebrafish
43	70.2	3.8	99395	2	AC010446 Homo sapi
44	70	3.8	125020	9	AF429315 Homo sapi
45	68.6	3.7	208878	2	AC141571 Rattus no

ALIGNMENTS

RESULT 1	AY061883	1786 bp	mRNA	linear	PRI 26-DEC-2001
LOCUS	AY061883				
DEFINITION	Homo sapiens naked protein (NKD)		mRNA, complete cds.		
ACCESSION	AY061883				
VERSION	AY061883.1	GI:17978536			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1786)				
AUTHORS	Yan, D., Wiesmann, M., Rohan, M., Chan, V., Fan, L., Jefferson, A. B., Guo, L., Sakamoto, D., Caotien, R. H., Fuller, J. H., Reinhard, C., Garcia, P. D., Randazzo, F. M., Escobedo, J., Fan, L. W. J., and Williams, L. T.,				
TITLE	Elevated expression of axin2 and nkhd mRNA provides evidence that Wnt/beta-catenin signaling is activated in human colon tumors				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14973-14978 (2001)				
PUBMED	11752446				
REFERENCE	2 (bases 1 to 1786)				
AUTHORS	Rohan, M., Chan, V., Yan, D., Fan, L., W. J. and Williams, L. T.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA				
FEATURES	Location/Qualifiers				
source	1..1786				
gene	/organism="Homo sapiens"				
CDS	/db_xref="taxon:9606"				
	1..1786				
	/gene="NKD"				
	159..1571				
	/gene="NKD"				
	/note="HNKD"				
	/codon_start=1				
	/product="naked protein"				
	/protein_id="ALJ32374.1"				
	/db_xref="GI:17978537"				
	/translation="MGKLSKPAVCRRRESPEGSFAVSAWARKGIEWIGRORCPGGVSGPRLAGTIRSTRELVDVLRPLTSEEDDRLTVALPPEKDGSGSGEKKMERVSPGSKKOLKPELQCDVSMEDSRQEWTFLLYDDNNKGYRBDITSLHTIYEVVDSVNHSPSTSSKMLRYKLTVAADGSGSKSVLVNQADLOSARPRATKPRPE				

[illegible]

ORIGIN

Query Match	96.1%	Score 1786;	DB 9;	Length 1786;
Best Local Similarity	100.0%	Pred. No. 2.3e-313;		
Matches 1786;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	66	AGTGGGACCGGGGAGACGACGGCAGAGAGCGGTCCGGGCGCGGCTCGGGCTTCGGTCCG	125
Db	1	AGTGGGACCGGGGAGACGACGGCAGAGAGCGGTCCGGGCGCGGCTCGGGCTTCGGTCCG	60
QY	126	CTCGGGGGCTGCTTCGGAGGAGGAGAGCCAAAGGAGGCGCCAGGCGCCGCGGGCCGGAGCG	185
Db	61	CTCGGGGGGCTGCTTCGGAGGAGGAGAGCCAAAGGAGGCGCCAGGCGCCGCGGGCCGGAGCG	120
QY	186	CATGGCTTAAAGGAAGCTTCCGGGCGCGGACCCCAAGATGGGGAACTTCACTCCAAAGC	245
Db	121	CATGGCTTAAAGGAAGCTTCCGGGCGCGGACCCCAAGATGGGGAACTTCACTCCAAAGC	180
QY	246	CGGCGCGCTGTGCAAGCGCAGAGAGAGCCGGAAAGTGACAGCTTCGCGCTGAGCGCTG	305
Db	181	CGGCGCGCTGTGCAAGCGCAGAGAGAGCCGGAAAGTGACAGCTTCGCGCTGAGCGCTG	240
QY	306	CTTGAGGCTCGGAAAGGCATCGAGAGTGGATCGGAGACAGCGCTGCCGGAGCGGTGCT	365
Db	241	CTTGAGGCTCGGAAAGGCATCGAGAGTGGATCGGAGACAGCGCTGCCGGAGCGGTGCT	300
QY	366	CGGAGACCCCGACACGTGGGTGGCGGGCACTATAGGCGGAAAGCACCCGGGAGGCTCGGG	425
Db	301	CGGAGACCCCGACACGTGGGTGGCGGGCACTATAGGCGGAAAGCACCCGGGAGGCTCGGG	360
QY	426	GCGACGTGTGAGAGACACGCTCAAGCAGAGAGAGAGAGACGACTTTCGCGCTGAGAGTGG	485
Db	361	GCGACGTGTGAGAGACACGCTCAAGCAGAGAGAGAGAGAGACTTTCGCGCTGAGAGTGG	420
QY	486	CCCTGCTCTCTGAAGAACTAGACGGGCTGGGCGACGGAGATGAGAGAAATGAGAGAG	545
Db	421	CCCTGCTCTCTGAAGAACTAGACGGGCTGGGCGACGGAGATGAGAGAAATGAGAGAG	480
QY	546	TGAGCGAAACCCCTGGCCAGAGCTCTCAAGAAAGCAGCTGAATTGGAAGAGCTCCAGTGGCAG	605
Db	481	TGAGCGAAACCCCTGGCCAGAGCTCTCAAGAAAGCAGCTGAATTGGAAGAGCTCCAGTGGCAG	540
QY	606	TGTCATAGAGAGAGACAGCCGGCAGAGAGTGAACCTTCACCTGTATGACTTTGAACA	665
Db	541	TGTCATAGAGAGAGACAGCCGGCAGAGAGTGAACCTTCACCTGTATGACTTTGAACA	600
QY	666	ACGGCAAGGTACCCCGAGAGAGACATCACACACTTGTCTGCAACACATCTATAGAGTGGTGG	725
Db	601	ACGGCAAGGTACCCCGAGAGAGACATCACACACTTGTCTGCAACACATCTATAGAGTGGTGG	660
QY	726	ACTCCTCTGTCAACCACTCCCAACATCAACCAAGATCTCTCGGGTAAAGCTCACCGTGG	785
Db	661	ACTCCTCTGTCAACCACTCCCAACATCAACCAAGATCTCTCGGGTAAAGCTCACCGTGG	720
QY	786	CCCCCGATGGCAGCCAGACCAAGAGAGAGGCTTGTTCATTCAGGCTGACCTTGCAAGCG	845
Db	721	CCCCCGATGGCAGCCAGACCAAGAGAGAGGCTTGTTCATTCAGGCTGACCTTGCAAGCG	780
QY	846	CAAAGCCCCCGAGCAGAGACCAAGCCCACTGAGAGACTGGCGAGCTGGGAGAAAGAACGAC	905
Db	781	CAAAGCCCCCGAGCAGAGACCAAGCCCACTGAGAGACTGGCGAGCTGGGAGAAAGAACGAC	840
QY	906	GAGCCCGGCTCAGGTTCCAGGGTGAACGCCCTCGAGACAGTCTTGAGCTACCACTT	965
Db	841	GAGCCCGGCTCAGGTTCCAGGGTGAACGCCCTCGAGACAGTCTTGAGCTACCACTT	900
QY	966	GCGTAGATAGACATCGAGAGAGAAACAATACTTAGATCTGGCCGGGATAGAAAATCT	1025

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgasbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov AKhter,N., Ayele,K., Beckerson-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dieterich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karllins,B., Kwong,P., Latic,P., Legaspi,R., Mahuro,Q.L., Mastello,C., Maesxrl,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Skantirlop,S., Thomas,P.J., Touchman,J.W., Tarragon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
JOURNAL	Alausner,R.D., Collins,F.S., Wagner,L., Shemmer,C.M., Schuler,G.D., Klausner,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Datchenko,L., Maruina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Udell,T.B., Toshitsuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abrahamson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shcherchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,C., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalka,U., Smalins,D.E., Scherch,A.J., Schein,J.E., Jones,S.J., and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences <i>Proc. Natl. Acad. Sci. U.S.A.</i> 99 (26), 16899-16903 (2002)
REFERENCE	12477932 2 (bases 1 to 2604)
AUTHORS	Strausberg,R.
JOURNAL	Direct Submission Submitted (25-Apr-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
COMMENT	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov Series: IRAL Plate: 47 Row: 0 Column: 23 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20127660
FEATURES	Location/Qualifiers
SOURCE	1. 2604 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:59666 IMAGE:664089" /tissue_type="Ovary, teratocarcinoma" /clone_1ib="NIH MGC_109" /lab_host="DH10B-R" /note="vector: pOTB7" 1. 2604 /gene="NKD1" /db_xref="LOCusID:85407" /db_xref="MIM:607851" 225. 1637 /gene="NKD1" /codon_start=1 /product="naked cuticle homolog 1" /protein_id="AAH51286.1" /db_xref="GI:30410966"
gene	
CDS	

Query Match	94.8%	Score 1763;	DB 9;	Length 2604;
Beet Local Similarity	100.0%	Pred. No. 1.9e-309;		
Matches 1763;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
66	AGTCGGGCGCGGCGACGCGCGGACGAGACCGGCTCCCGGCGCGCTCGGGGCTCCGGCTCGG			125
67	AGTCGGGCGCGGCGGCGACGCGCGGACGAGACCGGCTCCCGGCGCGCTCGGGGCTCCGGCTCGG			126
126	CTCGGGGGCTGCTTCGGGAGGAGGAGGACCAAGGAGGCGCCGAGGCGCCGGGCGGGGCGG			185
127	CTCGGGGGCTGCTTCGGGAGGAGGAGGACCAAGGAGGCGCCGAGGCGCCGGGCGGGGCGG			186
186	CATGGCTTAGGAGCGCTCCCGGCGCGCGGACCCGACATGGGGAACTTCACTTCGACG			245
187	CATGGCTTAGGAGCGCTCCCGGCGCGCGGACCCGACATGGGGAACTTCACTTCGACG			246
246	CGGCGCGCGGTGCGAAGGCGGAGGCGGAGACCCGGGAAAGTGACAGCTTCGCGCGGACGCGCTG			305
247	CGGCGCGCGGTGCGAAGGCGGAGGAGACCCGGGAAAGTGACAGCTTCGCGCGGACGCGCTG			306
306	CCTGGGCTCGGAGAGGCGATCGAGAGTGATGCGGAGACAGCGCTCCCGGCGGGTGCTT			365
307	CCTGGGCTCGGAGAGGCGATCGAGAGTGATGCGGAGACAGCGCTCCCGGCGGGTGCTT			366
366	CGGAGCCCCGACAGCTGCGGTTGGCGGGGACCATAGCGCGGAAAGCACCCGGGAGCTTCGTGG			425
367	CGGAGCCCCGACAGCTGCGGTTGGCGGGGACCATAGCGCGGAAAGCACCCGGGAGCTTCGTGG			426
426	GCGACGTGTGAGAGACACGCTCAGCGGAGGAAAGGAGAGAGAGACTTTCGGGCTGGAAGTGG			485
427	GCGACGTGTGAGAGACACGCTCAGCGGAGGAAAGGAGAGAGAGACTTTCGGGCTGGAAGTGG			486
486	CCCTGCGCTCTGAGAGAGACTGACGGGCTGGGCGACGGAGATGAGAGAGAAATGAGAGAG			545
487	CCCTGCGCTCTGAGAGAGACTGACGGGCTGGGCGACGGAGATGAGAGAGAAATGAGAGAG			546
546	TGAGCGAACCTGCGGAGGCTTCCAGAAAGAGCTGAAGTTTGAAGAGCTTCAAGTCGACG			605
547	TGAGCGAACCTGCGGAGGCTTCCAGAAAGAGCTGAAGTTTGAAGAGCTTCCAGTCGACG			606
606	TGTCATGAGAGAGAGACAGCGCGGACGAGGTGACCTTCAACCTGTATGACTTTGACAAACA			665
607	TGTCATGAGAGAGAGACAGCGCGGACGAGGTGACCTTCAACCTGTATGACTTTGACAAACA			666
666	ACGGCAAGGTCAACCCGAGAGAGACATCAACGACTTGTCTGACACACCATCTATGAGTGTGG			725
667	ACGGCAAGGTCAACCCGAGAGAGACATCAACGACTTGTCTGACACACCATCTATGAGTGTGG			726
726	ACTCTCTGTCAACCACTCCCAACATTCAGCAAGATGCTCGGGGTAAAGCTCACCGTGG			785
727	ACTCTCTGTCAACCACTCCCAACATTCAGCAAGATGCTCGGGGTAAAGCTCACCGTGG			786
786	CCCCCGATGCGAGCCAGGACAAAGAGAGCGTCTTGTGTAATTCAGGCTGACCTGCGAGAGCG			845
787	CCCCCGATGCGAGCCAGGACAAAGAGAGCGTCTTGTGTAATTCAGGCTGACCTGCGAGAGCG			846
846	CAAGGCCCCGAGCAGAGACCAAGCCCATTTGAGAGCTTCGGAGGAGGAGAAAGACAGC			905
847	CAAGGCCCCGAGCAGAGACCAAGCCCATTTGAGAGCTTCGGAGGAGGAGAAAGACAGC			906

```
QY 906 GAGCCCGCTCAGGTCAGGGGTGACAGCCGCTGAGACAGTCTGAGTCTACCAACCATY 965
Db 907 GAGCCCGCTCAGGTCAGGGGTGACAGCCGCTGAGACAGTCTGAGTCTACCAACCATY 966
QY 966 GCGTAGATGAGAACATCGAGAGAGAAACCACTAATATCTGCGGGATAGAAAACT 1025
Db 967 GCGTAGATGAGAACATCGAGAGAGAAACCACTAATATCTGCGGGATAGAAAACT 1026
QY 1026 ACAGAGTCCCAATTTGGGCTGGCTCCCTTCCGTCGGCCCAAGATCGAATCGCCCCC 1085
Db 1027 ACAGAGTCCCAATTTGGGCTGGCTCCCTTCCGTCGGCCCAAGATCGAATCGCCCCC 1086
QY 1086 GCACCTCCAAATCCCACTGATCTGCTCCATGAGCGGAGACCATCCACACACC 1145
Db 1087 GCACCTCCAAATCCCACTGATCTGCTCCATGAGCGGAGACCATCCACACACC 1146
QY 1146 GAAAGCCCCAAGGCGTGAGACCCGCTCTCTTCACTTCTTGACACCCCAATGCGCAAG 1205
Db 1147 GAAAGCCCCAAGGCGTGAGACCCGCTCTCTTCACTTCTTGACACCCCAATGCGCAAG 1206
QY 1206 TCTCAGAGCTTCAGACAAAGGCTCCGGGGCAACCCAGAGACGACACTTGTGAGT 1265
Db 1207 TCTCAGAGCTTCAGACAAAGGCTCCGGGGCAACCCAGAGACGACACTTGTGAGT 1266
QY 1266 CCCCCAAGGCCAGAGGCAAGAGTGTGGGTGTGGGCAAGTGTGCGCAGAGGGCAAGAAACA 1325
Db 1267 CCCCCAAGGCCAGAGGCAAGAGTGTGGGTGTGGGCAAGTGTGCGCAGAGGGCAAGAAACA 1326
QY 1326 AGCCCCCTTGGGACCCGCGCATTCCTGTGGGTGTCCTCCGACCTGTGCTCCAGCC 1385
Db 1327 AGCCCCCTTGGGACCCGCGCATTCCTGTGGGTGTCCTCCGACCTGTGCTCCAGCC 1386
QY 1386 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
Db 1387 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1446
QY 1446 AGAGCCAGAGAGGCTGCGGGGGCTGAGAGGACCACTGAGTGTGGGCTGTCTCTG 1505
Db 1447 AGAGCCAGAGAGGCTGCGGGGGCTGAGAGGACCACTGAGTGTGGGCTGTCTCTG 1506
QY 1506 GCGGGAGAGCACTGCGGGAGCTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1565
Db 1507 GCGGGAGAGCACTGCGGGAGCTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1566
QY 1566 CCGTTCAGAGACATGAGACCAACCAACCACTGAAACATCACACCATTTACCACTTCT 1625
Db 1567 CCGTTCAGAGACATGAGACCAACCAACCACTGAAACATCACACCATTTACCACTTCT 1626
QY 1626 ACCAGACATAGAGCCCTCCCAAGGCCCCACCTGACATGTGAAGAGACCCACCCCGA 1685
Db 1627 ACCAGACATAGAGCCCTCCCAAGGCCCCACCTGACATGTGAAGAGACCCACCCCGA 1686
QY 1686 CACCAAGGCAATTAATTTATTAATTAATTAATTTATTTATTTATTTATTTATTAATA 1745
Db 1687 CACCAAGGCAATTAATTTATTAATTAATTTATTTATTTATTTATTTATTTATTAATA 1746
QY 1746 ATTATTTGTACTCCACTTAATTTAGAGCTCACTAGTGAAGAAATCTATGAAACACAG 1805
Db 1747 ATTATTTGTACTCCACTTAATTTAGAGCTCACTAGTGAAGAAATCTATGAAACACAG 1806
QY 1806 AACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1828
Db 1807 AACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1829
```

```
RESULT 3
LOCUS AB062886 1788 bp mRNA linear PRI 20-OCT-2001
DEFINITION Homo sapiens mRNA for Dvl-1-binding protein NKD1, complete cds.
ACCESSION AB062886
VERSION AB062886.1 GI:16303259
KEYWORDS
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Katoh,M.
TITLE Molecular cloning, gene structure, and expression analyses of NKD1
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PubMed 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh,M.
TITLE Direct Submision
JOURNAL Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center
Research Institute, Genetics and Cell Biology Section; Tsukiji
5-chome, Chuo-ku, Tokyo 104-0045, Japan [E-mail:mkatoh@ncc.go.jp,
Tel:81-3-3542-2511]
FEATURES
source
1..1788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="human fetal tissues"
1..1788
/gene="NKD1"
6..1418
/gene="NKD1"
/codon_start=1
/product="Dvl-1-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16303260"
/translation="MGKHSKPAVCKKRESPEGSDPAVSAAMARKGIEIGRORCP
GYSGRQLRLACTIGRSPELVDYLRDLSEEDDPLFALPEKTDIGSGDE
KMERVSEPCSKQKQKTEBELQDYMEDSDQENTFTLYIDNNKATREBITSL
HTIYEVNDSVNSPTSSKMLRVKLTVAIPGSSKSVLVNQADLQSAKRAETKPE
DLRSWEKQKAPLRFQDSRLQSGGCHVNDENIBRNHYDLAIGENTSGFGPS
PSVAQSELPFPTSPNPTSRSHPEPAIHIPKRPQGVDPASFPLDPIAKVSELQK
LRGTODGSKHFVSPKAGKSVGVGVARAKRPLGPAPIPAVPSAHLAASPLALP
SLAPLHKHKHAKESQDQCRGLQAPLASGCVLGRHRELPALVVTYSQAGPQV
RHHHHHHHHHHHHHHHHYQI"
```

```
ORIGIN
Query Match 86.6%; Score 1610; DB 9; Length 1788;
Best Local Similarity 100.0%; Pred. No.1e-281;
Matches 1610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 219 CCAAGCATGGGAAACTTCACTCCAGCGCGCGCTGTGCAAGCCGAGGAGAGCCCGG 278
Db 1 CCAAGCATGGGAAACTTCACTCCAGCGCGCGCGCTGTGCAAGCCGAGGAGAGCCCGG 60
QY 279 AAGGTGACACTTTCGCGTGAAGGCTGCTGGGCTCGGAAGGGCATTCAGAGATGATCG 338
Db 61 AAGGTGACACTTTCGCGTGAAGGCTGCTGGGCTCGGAAGGGCATTCAGAGATGATCG 120
QY 339 GGAAGACGCGCTCGCGGCGGCGGTCTCGGGAACCCCGACAGCTGCGGTTGGCGGACCA 398
Db 121 GGAAGACGCGCTCGCGGCGGCGGTCTCGGGAACCCCGACAGCTGCGGTTGGCGGACCA 180
QY 121 GGAAGACGCGCTCGCGGCGGCGGTCTCGGGAACCCCGACAGCTGCGGTTGGCGGACCA 180
Db 399 TAGCGCAAGACACCCCGGAGCTGTGGGCGACGTTGTGAGAGACAGCTCAGCGAGAG 458
QY 181 TAGCGCAAGACACCCCGGAGCTGTGGGCGACGTTGTGAGAGACAGCTCAGCGAGAG 240
Db 459 AGAGAGACGACTTTCGCGTGAAGGTGCTGCTCTGTGAAGAGCTGACGGGCTGGGCA 518
QY 241 AGAGAGACGACTTTCGCGTGAAGGTGCTGCTCTGTGAAGAGCTGACGGGCTGGGCA 300
Db 519 GCGAGATGAGAGAGATGAGAGATGAGAGATGAGAGACCTTCGACGCTCCAGAGAGAGC 578
QY 301 GCGAGATGAGAGAGATGAGAGATGAGAGATGAGAGACCTTCGACGCTCCAGAGAGAGC 360
Db 579 TGAAGTTTGAAGAGCTTCAGTGCAGCTGTCCATGAGAGAGAGACAGCCGCGCAGAGTGA 638
QY 361 TGAAGTTTGAAGAGCTTCAGTGCAGCTGTCCATGAGAGAGAGACAGCCGCGCAGAGTGA 420
```

QY	639	CCTTCACCCCTGTAAAGACCTTTGACAAACAAGGCAAGGTCAACCCGAGAGACATCAACACT	698
Db	421	CCTTCACCCCTGTAAAGACCTTTGACAAACAAGGCAAGGTCAACCCGAGAGACATCAACACT	480
QY	699	TGCTGCACACCATCTATAGAGGTGTGTGACTCCTCTGTCAACCACTCCCAACATCCAGCA	758
Db	481	TGCTGCACACCATCTATAGAGGTGTGTGACTCCTCTGTCAACCACTCCCAACATCCAGCA	540
QY	759	AGATGCTCGGGGTAAAGCTCAACCTGTGGCCCCGATGGCAAGCCAGACAAAGAGACCTCC	818
Db	541	AGATGCTCGGGGTAAAGCTCAACCTGTGGCCCCGATGGCAAGCCAGACAAAGAGACCTCC	600
QY	819	TTGTCAATCAGAGCTGCACTGCAAGACGCAAGGCCCGGACAGAGACCAAGCCCACTGAGG	878
Db	601	TTGTCAATCAGAGCTGCACTGCAAGACGCAAGGCCCGGACAGAGACCAAGCCCACTGAGG	660
QY	879	ACCTGCGGAGCTGGGAGAAAGACGCAAGCCCGCTCAGTTCAAGGTGTGACAGCCGC	938
Db	661	ACCTGCGGAGCTGGGAGAAAGACGCAAGCCCGCTCAGTTCAAGGTGTGACAGCCGC	720
QY	939	TGAGCAGATCTGGGCTGCTACCAACCTTGCGTATGATGAGAACATGAGAGAGAAACCACT	998
Db	721	TGAGCAGATCTGGGCTGCTACCAACCTTGCGTATGATGAGAACATGAGAGAGAAACCACT	780
QY	999	ACTTAGATCTGCGCGGGATAGAAAACTACAGTCCCAATTTGGGCTGGCTCCCTTCG	1058
Db	781	ACTTAGATCTGCGCGGGATAGAAAACTACAGTCCCAATTTGGGCTGGCTCCCTTCG	840
QY	1059	TGGGCCAGAAATTCAGAACTGCCCCCGGCACTCCCAATCCCACTGATCTGGCTCCATG	1118
Db	841	TGGGCCAGAAATTCAGAACTGCCCCCGGCACTCCCAATCCCACTGATCTGGCTCCATG	900
QY	1119	AGCGCGAAGCATCCACATCTCCCAACCGAAAGCCCAAGCGGTGAGCCGGGCTCTTCC	1178
Db	901	AGCGCGAAGCATCCACATCTCCCAACCGAAAGCCCAAGCGGTGAGCCGGGCTCTTCC	960
QY	1179	ACTTCTTTGACACCCCAATTCGCCAAGTCTCAAGCTTCAGCAACGCTCGGGGCAACC	1238
Db	961	ACTTCTTTGACACCCCAATTCGCCAAGTCTCAAGCTTCAGCAACGCTCGGGGCAACC	1020
QY	1239	AGGACGGGAGCAAGCACTTTGTAGGTCCCCCAAGGGCCCAAGGGCAAGGTGTGGTGG	1298
Db	1021	AGGACGGGAGCAAGCACTTTGTAGGTCCCCCAAGGGCCCAAGGTGTGGTGGTGG	1080
QY	1299	GCACAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGTGGGACCCGCACTCCCTGGGTGT	1358
Db	1081	GCACAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGTGGGACCCGCACTCCCTGGGTGT	1140
QY	1359	CCCCCTCGGCCACTGTGCTGCGAAGCCGGGCTCTCTCTCCCTTAAAGCCCTCGGGC	1418
Db	1141	CCCCCTCGGCCACTGTGCTGCGAAGCCGGGCTCTCTCTCCCTTAAAGCCCTCGGGC	1200
QY	1419	ACAAGAAGCAACAAGACCGAGGCCAAGAGAGCGCAGCAGGCTGCGGGGGCTGCAAGGAC	1478
Db	1201	ACAAGAAGCAACAAGACCGAGGCCAAGAGAGCGCAGCAGGCTGCGGGGGCTGCAAGGAC	1260
QY	1479	CACTGGCTCAAGTGGCCCTGTCTGTGGGGCGGAGACCTGCGGAGGCTGCGCGCTTGG	1538
Db	1261	CACTGGCTCAAGTGGCCCTGTCTGTGGGGCGGAGACCTGCGGAGGCTGCGCGCTTGG	1320
QY	1539	TGGGTATGAGAGCCAGGCGGGGACGGGGTCCAGAGCATTGAGAGCAACAACAACAATG	1598
Db	1321	TGGGTATGAGAGCCAGGCGGGGACGGGGTCCAGAGCATTGAGAGCAACAACAACAATG	1380
QY	1599	AACATCAACACATTACCAACCACTTTCAACAAGCATTAGAGCCCTCCCAAGGCCCAACC	1658
Db	1381	AACATCAACACATTACCAACCACTTTCAACAAGCATTAGAGCCCTCCCAAGGCCCAACC	1440
QY	1659	CTGCATATGAGAGACCCCAACCCCGAGAACCAACAAGGATTTATTTCTATTAATATATG	1718
Db	1441	CTGCATATGAGAGACCCCAACCCCGAGAACCAACAAGGATTTATTTCTATTAATATATG	1500

Oy	1719	TTATTATGATGATTATGGTATTATTAATAATTATGTACTCCAGTAATATTAGTAGCCT	1778
Dd	1501	TTATTATGATGATTATGGTATTATTAATAATTATGTACTCCAGTAATATTAGTAGCCT	1560
Oy	1779	ACATGTAGAAGATCTATGAAACAAGAACTAACTTTATTATTATATGTT	1828
Dd	1561	ACATGTAGAAGATCTATGAAACAAGAACTAACTTTATTATTATATGTT	1610
<hr/>			
RESULT 4 AF358135 1438 bp mRNA linear PRI 29-MAY-2001			
LOCUS AF358135 Homo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.			
DEFINITION Homo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.			
ACCESSION AF358135			
VERSION AF358135.1 GI:14211713			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1. 1438			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/chromosome="16"			
/map="16q12"			
1. .1438			
/gene="NKD1"			
8. .1420			
/gene="NKD1"			
/note="contains EFx domain; similar to Drosophila melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"			
/codon_start=1			
/product="naked cuticle-1"			
/protein_id="AAK57484.1"			
/db_xref="GI:14211714"			
/translation="MGKLHKRPAVAVCRRESPEQSPRAVSAAARKGIEIMIRORCP			
GGYSQPOLRLAGTITGRSTRELDVDVLRTLSSEEDDFLEVALPREKTIDGSGBE			
KKMERVSEPQSGSKQLKFELQCDVMSDEDSQEWTFTLYDDNNKVTRBETISLL			
HTIYEVSDSVNHSPTSSKMRLRYLTVAIPDGSGSKRVLVNQDLQAPRAETKPTP			
DLSWEKKORALPLRFQGDSDLQSGCYCHNCVDENIERRNHYLDLAGIENTSQFGPS			
PSVAKGSLPRTSNPTFRSRSHBEBAHIHRKPOGVDPASPHFLDRPIKVESLQOR			
LRGQDSKHFIVSPKAGSGSYGVGHVARANKPKPGAPIPAVSPAHIAASPALLP			
SLAPLGKKHKHRAKESQGCGRCGLQAPLASGYVLGRHLRELPAVVYESQAGCPVQ			
RHEHHHHHHHHHPYTQT"			
<hr/>			
ORIGIN			
Query Match 77.4%; Score 1438; DB 9; Length 1438;			
Best Local Similarity 100.0%; Pred. No. 1,6e-250;			
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Oy	217	CCCCAGATGGGGAACCTTCACCTCAAGCCGCCGCCTGTGCAGAGCGAAGGAGACC	276
Dd	1	CCCCAGATGGGGAACCTTCACCTCAAGCCGCCGCCTGTGCAGAGCGAAGGAGACC	60
Oy	277	GGAAGCGTACAGCTTGCCCGCTGAGCGCTGCTGCTCGAGAGGCGCATCGAGAGTGAT	336
Dd	61	GGAAGCGTACAGCTTGCCCGCTGAGCGCTGCTGCTCGAGAGGCGCATCGAGAGTGAT	120

OY	337	CGGGAGACAGCGCTGCCCGGGCGGTGTCTCGGGACCCCGACACCGTGGGTTGGGGGCAC	396
Db	121	CGGAGACAGGGCTGCCCGGGCGGTGTCTCGGGACCCCGACAGCTGGGTTGGGGGCAC	180
OY	397	CATAGGGCCGAAGCACCCGGAGACTCGTAGGGGAGAGGTGTGGAGACACGCTCAGCGAGGA	456
Db	181	CATAGGCCGAAGCACCCGGAGACTCGTAGGGGAGAGGTGTGGAGACACGCTCAGCGAGGA	240
OY	457	AGAGAGGACGACCTTTCGGCTGGAGAGTGGCCCTGCTCTGAGGAAGACTGACGGGCTGGG	516
Db	241	AGAGAGGACGACCTTTCGGCTGGAGAGTGGCCCTGCTCTGAGGAAGACTGACGGGCTGGG	300
OY	517	CAGCGGAGATGAGAAAGATGAGAGAGTGAAGGAAACCTTCGCCAGGCTTCAGAAAGCA	576
Db	301	CAGCGGAGATGAGAAAGATGAGAGAGTGAAGGAAACCTTCGCCAGGCTTCAGAAAGCA	360
OY	577	GCTGAAGTTTAAAGAGCTCCAGTGGGACGTGTCCATGGAGGAGGACAGCGGGAGAGTG	636
Db	361	GCTGAAGTTTAAAGAGCTCCAGTGGGAGGTGTCCATGGAGGAGGACAGCGGGAGAGTG	420
OY	637	GACCTTCACCCCTGTATGACTTTTGAACAACAACGGCAAGGTCAACCGAGAGGACATCACAG	696
Db	421	GACCTTCACCCCTGTATGACTTTTGAACAACAACGGCAAGGTCAACCGAGAGGACATCACAG	480
OY	697	CTTGCTGACACCACTATATGAGGTGTGACTCTCTGTCAACACTTCCCAATCCAG	756
Db	481	CTTGCTGACACCACTATATGAGGTGTGACTCTCTGTCAACACTTCCCAATCCAG	540
OY	757	CAAGAATGCGGGGTTAAAGCTCACCGTAGCCCCCGGATGGCGAGCCAGACAAAGAAAGGT	816
Db	541	CAAGAATGCGGGGTTAAAGCTCACCGTAGCCCCCGGATGGCGAGCCAGACAAAGAAAGGT	600
OY	817	CCTTGTCATAGGCTGACCTGACAGCGCAAGGCCCCGAGCAGAGACCAAGCCCACTGA	876
Db	601	CCTTGTCATAGGCTGACCTGACAGCGCAAGGCCCCGAGCAGAGACCAAGCCCACTGA	660
OY	877	GGACCTGCGGAGCTGGGAGAAAGAACGACGAGCCCCGCTCAGGTTCCAGGGTGAACGCCG	936
Db	661	GGACCTGCGGAGCTGGGAGAAAGAACGACGAGCCCCGCTCAGGTTCCAGGGTGAACGCCG	720
OY	937	CCTGAGAGAGTCTGGCTGTACTCAACATTTGGGTATATGAAACATCCAGAGAGAAACCA	996
Db	721	CCTGAGAGAGTCTGGCTGTACTCAACATTTGGGTATATGAAACATCCAGAGAGAAACCA	780
OY	997	CTACTTAGATCTCGCGCGGATAGAAAACTACACGTCCCAATTTGAGGCTCGGCTCCCTTC	1056
Db	781	CTACTTAGATCTCGCGCGGATAGAAAACTACACGTCCCAATTTGAGGCTCGGCTCCCTTC	840
OY	1057	CGTGGCCGAGAGTCAGAACTGCCGCCCGGACCTCCATCCCACTCGATCTGCTCCCA	1116
Db	841	CGTGGCCGAGAGTCAGAACTGCCGCCCGGACCTCCATCCCACTCGATCTGCTCCCA	900
OY	1117	TGAGCCGGAAAGCCATCCCATTCACACCCGAAAGCCCAAGCGTGAACCCGGCTCCTT	1176
Db	901	TGAGCCGGAAAGCCATCCCATTCACACCCGAAAGCCCAAGCGTGAACCCGGCTCCTT	960
OY	1177	CCACTTCTTGTGACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGAGCTCCGGGAGCA	1236
Db	961	CCACTTCTTGTGACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGAGCTCCGGGAGCA	1020
OY	1237	CCAGGACGGGAGCAGCACTTTGTAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTG	1296
Db	1021	CCAGGACGGGAGCAGCACTTTGTAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTG	1080
OY	1297	GGGCGACGTGGCAGAGGGGCGCAAGAAACAAGCCCCCTTGGGACCCGCAATCCCTGGGT	1356
Db	1081	GGGCGACGTGGCAGAGGGGCGCAAGAAACAAGCCCCCTTGGGACCCGCAATCCCTGGGT	1140
OY	1357	GTCCCCCTCCGGCCCACTGTGGCTGCAGGCGGGCCCTCTCCCTCCCAAGCCCCCTCGG	1416
Db	1141	GTCCCCCTCCGGCCCACTGTGGCTGCAGGCGGGCCCTCTCCCTCCCAAGCCCCCTCGG	1200
OY	1417	GCACAAAGAGCACAAAGCACCGAGCCCAAGAGAGCCAGAGGAGTGCCTGGGCTGCAGGC	1476

Db	1201	GCACAGAGACACAGACCGAGCCAGAGAGAGCCAGCAGGGGCTGCGGGCCCTGCAGGC	1260
QY	1477	ACCACTGGCCCTCAGGTGGCCCTGTCTCTGGGCGGGAGCACCTGCGGAGCTGCCCCCTT	1536
Db	1261	ACCACTGGCCCTCAGGTGGCCCTGTCTCTGGGCGGGAGCACCTGCGGAGCTGCCCCCTT	1320
QY	1537	GGTGTGTATATAGAGCCAGGCGCGGGCAGCGGCTCCAGAGCATATAGACACCAACCAACCA	1596
Db	1321	GGTGTGTATATAGAGCCAGGCGCGGGCAGCGGCTCCAGAGCATATAGACACCAACCAACCA	1380
QY	1597	TGAACATCACCACCATTAACCACTTCTACACACATAGAGCCCTCCACGGAGCC	1654
Db	1381	TGAACATCACCACCATTAACCACTTCTACACACATAGAGCCCTCCACGGAGCC	1438
RESULT 5			
AR405917	AR405917	1438 bp	DNA
LOCUS	Sequence 5 from patent US 6630323.	Linear	PAT 18-DEC-2003
DEFINITION	AR405917		
ACCESSION	AR405917		
VERSION	AR405917.1	GI:40154965	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1438)		
AUTHORS	Scott, M., Zeng, W. and Wharton, K.		
TITLE	Naked cuticle genes and their uses		
JOURNAL	Patent: US 6630323-A 5 07-Oct-2003;		
FEATURES	Location/Qualifiers		
source	1..1438		
ORIGIN	/organism="unknown"		
	/mol_type="genomic DNA"		
Query Match	77.2%;	Score 1434.8;	DB 6; Length 1438;
Best Local Similarity	99.9%;	Pred. No. 6.1e-250;	
Matches 1436;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
QY	217	CCCCAGCATGGGGAACTTCACTCCAGCGCGCGCGCTGTGCAAGCGCAGAGGAGGCC	276
Db	1	CCCCAGCATGGGGAACTTCACTCCAGCGCGCGCGCTGTGCAAGCGCAGAGGAGGCC	60
QY	277	GGAAGGTGACAGCTTCGCGCTGAGCGCTGCTCGGAGGCGATCGAGAGTGGAT	336
Db	61	GGAAGGTGACAGCTTCGCGCTGAGCGCTGCTCGGAGGCGATCGAGAGTGGAT	120
QY	337	CGGAGACAGCGCTGCCCGGGCGGTGTCTCGGACCCCGACAGCTGCGGTTGCGGGTAC	396
Db	121	CGGAGACAGCGCTGCCCGGGCGGTGTCTCGGACCCCGACAGCTGCGGTTGCGGGTAC	180
QY	397	CATAGGCCGAGACACCGCGGAGCTCTGTGGGGAGCGGTGTGAGACACGCTCAGCGAGGA	456
Db	181	CATAGGCCGAGACACCGCGGAGCTCTGTGGGGAGCGGTGTGAGACACGCTCAGCGAGGA	240
QY	457	AGAGAGGACACATTTGGCTGGAAGTGGCCCTCTCTGAGGAAGACTGACGCGCTGGG	516
Db	241	AGAGAGGACACATTTGGCTGGAAGTGGCCCTCTCTGAGGAAGACTGACGCGCTGGG	300
QY	517	CAGCGGATGAGAAAGATGATGAGAGGTAGCGAACCTTGCCCGAGCTTCAAGAAACA	576
Db	301	CAGCGGATGAGAAAGATGATGAGAGGTAGCGAACCTTGCCCGAGCTTCAAGAAACA	360
QY	577	GCTGAAGTTTGAAGCTCCAGTCCGACGTCTCATATGAGGAGAGCAGCGCGGAGGTG	636
Db	361	GCTGAAGTTTGAAGCTCCAGTCCGACGTCTCATATGAGGAGAGCAGCGCGGAGGTG	420
QY	637	GACCTTCACTCTGATGACTTTTGAACAACAAGGTCACCCGAGAGGACATCAACAG	696
Db	421	GACCTTCACTCTGATGACTTTTGAACAACAAGGTCACCCGAGAGGACATCAACAG	480
QY	697	CTTGTGACACCATATATAGAGTGTGTGACTCTCTGTCAACCACTCCCAACATCCAG	756

```

Db      481  CTTGCTGCACACCATATATAGAGTGTGGAGCTCTCTGTCAACCACTCCCAACATCCAG 540
Qy      757  CAAGATGCTGGGGTAAAGCTCAACCGTGGCCCCGATGGCAGCCAGAGCAAGAGAGCGT 816
Db      541  CAGATGCTGGGGTAAAGCTCAACCGTGGCCCCGATGGCAGCCAGAGCAAGAGAGCGT 600
Qy      817  CCTGTTCATCAGAGCTGACCTGCAAGAGCCGAAAGGCCCGGAGAGCAAGCCCACTGA 876
Db      601  CCTGTTCATCAGAGCTGACCTGCAAGAGCCGAAAGGCCCGGAGAGCAAGCCCACTGA 660
Qy      877  GAGCTGCGGAGCTGGGAGAAAGAGCAGAGCCCGCTCAGAGTTCCAGGGGTGACAGCG 936
Db      661  GAGCTGCGGAGCTGGGAGAAAGAGCAGAGCCCGCTCAGAGTTCCAGGGGTGACAGCG 720
Qy      937  CCTGAGCAGCTGCTGCTACCAACATTTGCGATGAGAGCAATCGAGAGAGAAACA 996
Db      721  CCTGAGCAGCTGCTGCTACCAACATTTGCGATGAGAGCAATCGAGAGAGAAACA 780
Qy      997  CTACTTATGATCTGCGCGGATAGAAAATCAAGTCCCAATTTGGGCTTGGCTTCCCTTC 1056
Db      781  CTACTTATGATCTGCGCGGATAGAAAATCAAGTCCCAATTTGGGCTTGGCTTCCCTTC 840
Qy      1057  CGTGGCCGAGAGTCAAGACTGGCCCCCGACCTCCCAATCCCACTGATCTGCTCCCA 1116
Db      841  CGTGGCCGAGAGTCAAGACTGGCCCCCGACCTCCCAATCCCACTGATCTGCTCCCA 900
Qy      1117  TGAAGCCGAGAGTCAAGACTGGCCCCCGACCTCCCAATCCCACTGATCTGCTCCCA 1176
Db      901  TGAAGCCGAGAGTCAAGACTGGCCCCCGACCTCCCAATCCCACTGATCTGCTCCCA 960
Qy      1177  CCACTTCTTGAACACCCCAATTCGCAAGGCTCTCAGAGCTCCAGCAAGGCTCCGGGGCAC 1236
Db      961  CCACTTCTTGAACACCCCAATTCGCAAGGCTCTCAGAGCTCCAGCAAGGCTCCGGGGCAC 1020
Qy      1237  CCAGAGCGGAGAGCAAGCACTTTGTAGAGTCCCAAGGCGCCAGAGGCAAGAGTGGGTGT 1296
Db      1021  TCAGAGCGGAGAGCAAGCACTTTGTAGAGTCCCAAGGCGCCAGAGGCAAGAGTGGGTGT 1080
Qy      1297  GGGGCAAGTGGCAGAGGGGCAAGAAACAGAGCCCTCTTGGGAGCCCGCATCTCTGGCGT 1356
Db      1081  GGGGCAAGTGGCAGAGGGGCAAGAAACAGAGCCCTCTTGGGAGCCCGCATCTCTGGCGT 1140
Qy      1357  GTTCCCTTCCGCCCACTGCTGCGAGCCCGGCGCTCTCTCCCTCTAGCCCTCTGG 1416
Db      1141  GTTCCCTTCCGCCCACTGCTGCGAGCCCGGCGCTCTCTCCCTCTAGCCCTCTGG 1200
Qy      1417  GCAAGAGAGCAAGAGCAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
Db      1201  GCAAGAGAGCAAGAGCAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy      1477  ACCACTGAGCTCAGTGGCCCTGTCTGAGGGCGGAGACCTGCGGAGAGTGGCCGCTT 1536
Db      1261  ACCACTGAGCTCAGTGGCCCTGTCTGAGGGCGGAGAGACCTGCGGAGAGTGGCCGCTT 1320
Qy      1537  GGTGTGTATGAGAGCAGAGCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
Db      1321  GGTGTGTATGAGAGCAGAGCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy      1597  TGAACATCAACCACTTACCAACATTTCTACAGAGCATAGAGCCCTCCAGAGAGAGAG 1654
Db      1381  TGAACATCAACCACTTACCAACATTTCTACAGAGCATAGAGCCCTCCAGAGAGAGAG 1438

```

```

RESULT 6
BC034838      1765 bp      mRNA      linear      ROD 29-JUN-2004
LOCUS         BC034838
DEFINITION    Mus musculus naked cuticle 1 homolog (Drosophila), mRNA (cDNA clone
ACCESSION     BC034838
VERSION       BC034838.1 GI:22028144
KEYWORDS      MGC.
SOURCE        Mus musculus (house mouse)

```

```

ORGANISM      Mus musculus
REFERENCE     1 (bases 1 to 1765)
AUTHORS      Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
              Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
              Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
              Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
              Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
              Stapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L.,
              Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
              Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
              Abramson, R.D., Mulhany, S.J., Bosak, S.A., McSwan, P.J.,
              McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
              Worrley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W.,
              Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
              Fahey, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S.,
              Sanchez, A., Schein, J.E., Jones, S.J., Jones, S.J., and Marra, M.A.
              Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
              Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
              Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E.,
              Scherch, A., Schein, J.E., Jones, S.J., Jones, S.J., and Marra, M.A.
              Generation and initial analysis of more than 15,000 full-length
              human and mouse cDNA sequences
              Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)
              12477932
              2 (bases 1 to 1765)
              Strausberg, R.
              Direct Submission
              Submitted (29-JUL-2002) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genome Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
              USA
              NIH-MGC Project URL: http://mgc.nci.nih.gov
              Contact: MGC help desk
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: Marcello Bento Soares, Ph.D.
              CDNA Library Preparation: M. Bento Soares, University of Iowa
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLB)
              DNA Sequencing by: Genome Sequence Centre,
              BC Cancer Agency, Vancouver, BC, Canada
              info@cgsc.bc.ca
              Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butlerfield,
              Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth
              Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
              Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
              Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu,
              Parvaneh Saeedi, JR Santos, Angelique Scherch, Ursula Skalska,
              Duane Smalins, Jeff Scott, Miranda Tsai, George Yang, Jacquele
              Schein, Asim Siddiqui, Rob Holt, Marco Marra.
              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LNLB at: http://image.jhmi.gov
              Series: IRAP Plate: 67 Row: 9 Column: 5
              This clone was selected for full length sequencing because it
              passed the following selection criteria: matched mRNA gi: 14456707.
              Location/Qualifiers
                1. 1765
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="MGC:40995 IMAGE:1429943"
                  /tissue_type="mammary gland"
                  /clone_lib="Soares_mammary_gland_NbMWG"
                  /lab_host="DH10B"
                  /note="Vector: pT73-Pac"
                1. 1765
                  /gene="Nkd1"
                  /db_xref="LocustID:93960"
                  /db_xref="MGI:2135954"
                138..1553
                  /gene="Nkd1"

```

```
/codon_start=1
/product="naked cuticle 1 homolog"
/protein_id="AAH34838.1"
/db_xref="GI:22028145"
/db_xref="MGJ:2135954"
/translation="MGKHSKPAAVCKRRESPEGDSPAVSAAMARKIEIMIGRORCP
GSVSGPROLALAGTGTRELVGDTREALGREDDEPFLVLAPEKIDISGSDPE
KMRERLSEPGQASKKOLKRELEODVSEEDSRBFTLYDPDNKRYTRBIDTSLI
HTIEVDSSVNSPTSKTLRYKLVAFPGSQSKSVLNFHIDLSQTRPADIKPAE
ELRGWKQKQAPLRFQDSDLQPDCAHNCVDENIERNNHYLDLAGIENTSOFGPS
PSVAOKSELPRISNPTRSRSHPEBAHVBRPOQDPSFHLIDPFAKASLQOR
LRGTODSKHFRVSPKQAGKMGKHGARARSPPLVPTTHYVSPAHLATBPALP
TLAFLGHKKHKHAKESQASCRGLQGLAAGSTVMRBEQVRELPAVVVYESQAGQAV
ORHHHHHHHHHHHYFPYQ"
```

ORIGIN

Query Match 67.8%; Score 1259.6; DB 10; Length 1765;
Best Local Similarity 84.3%; Pred. No. 3.6e-218;
Matches 1497; Conservative 0; Mismatches 254; Indels 25; Gaps 6;

```
QY CGACGCGCGACGAGCGCGTCCGCGCGCGCGCGCTCGGCTCGGCTCGGCGGCTGCT 138
Db 1 CGACGCGCGACGAGCGCGTCCGCGCGCGCGCGCTCGGCTCGGCTCGGCGGCTGCT-- 58
QY 139 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 198
Db 59 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 116
QY 199 CGCTCCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 258
Db 117 GCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 172
QY 259 CAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
Db 173 CAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
QY 319 GGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 378
Db 233 AGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 292
QY 379 GCTGCGGTTGGCGGCGACCATAGGCGCGAAGCAGCGGAGGCTCGTGGGCGAGCTGTTAG 438
Db 293 GCTGAGATTGGCAGGCGACTGTGGTCAAGAGCACTCGGGAATCGTGGGTGACATTTCTAG 352
QY 439 AGACACGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 498
Db 353 AGAGGCTCTCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
QY 499 GAAGACTGAGCGGCTGGGCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558
Db 413 GAAGATTGACAGCTTAGGTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472
QY 559 CCCAGGCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 618
Db 473 CAGAGGCTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532
QY 619 GAGACGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 678
Db 533 GAGACGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
QY 679 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 738
Db 593 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
QY 739 CCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 798
Db 653 CCAATTCCTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 712
QY 799 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 858
Db 713 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 772
```

```
QY 859 AGAGACCAAGCCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 918
Db 773 AGACACCAAGCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 832
QY 919 GTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 978
Db 833 GTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 892
QY 979 CATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1038
Db 893 CATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 952
QY 1039 TGGAGCTGAGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1098
Db 953 TGAAGCCGAGATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1012
QY 1099 CACTCGATCTCGTCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
Db 1013 CACTCGCTCTCGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1072
QY 1159 CGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1218
Db 1073 TGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1132
QY 1219 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1278
Db 1133 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1192
QY 1279 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1338
Db 1193 GAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1252
QY 1339 ACCGCGCATCCCTCGGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1398
Db 1253 ACCGCGCATCCCTCGGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1312
QY 1399 CTCCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
Db 1313 CACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1372
QY 1459 CTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515
Db 1373 CTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1432
QY 1516 CTTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1575
Db 1433 GGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1492
QY 1576 ACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1635
Db 1493 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1552
QY 1636 GAGCGGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695
Db 1553 GAGCGGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1598
QY 1696 CATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1755
Db 1599 CATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1658
QY 1756 CTCACATTAATTTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 1815
Db 1659 CTCACATTAATTTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 1718
QY 1816 TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1851
Db 1719 TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1754
```

RESULT 7
AR405916
LOCUS
DEFINITION Sequence 3 from patent US 6630323. 1731 bp. DNA linear PAT 18-DEC-2003

ACCESSION AR405916
 VERSION AR405916.1 GI:40154964
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 1731)
 AUTHORS Scott, M., Zeng, W. and Wharton, K.
 TITLE Naked cuticle genes and their uses
 JOURNAL Patent: US 6630323-A 3 07-OCT-2003;
 FEATURES
 location/Qualifiers
 source 1..1731
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Query Match 65.7%; Score 1221.4; DB 6; Length 1731;
 Best Local Similarity 83.9%; Pred. No. 3.1e-211;
 Matches 1460; Conservative 0; Mismatches 256; Indels 25; Gaps 6;

QY 82 CCGCGGAGAGGCGCTCCGCGCGCGCTCGAGCTCCGCTCGAGCTCGAGGAGCTTGC 141
 Db 6 CCGGCGCGAGGCGCTCCGCGCGCGCTTAAGCTCGCTCGAGCGCGGACTG--TGA 63
 QY 142 GAG 201
 Db 64 GAG 121
 QY 202 TCCGCGCGCGCGAG 261
 Db 122 CCGCGCGCGC---GCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
 QY 262 GCGCAGGAG 321
 Db 178 GCGCAGGAG 237
 QY 322 CATCGAG 381
 Db 238 CATCGAG 297
 QY 382 GCGGTTGGCGGAG 441
 Db 298 GAGATTGGAG 357
 QY 442 CAGCTCAG 501
 Db 358 GGTCTCGTGAAG 417
 QY 502 GACTGAG 561
 Db 418 GATCGAG 477
 QY 562 AGGCTCCAG 621
 Db 478 GGGCTCCAG 537
 QY 622 CAGCGGAG 681
 Db 538 CAGCGGAG 597
 QY 682 AG 741
 Db 598 TGAAG 657
 QY 742 CTCCCAAG 801
 Db 658 TTCCCAAG 717
 QY 802 GAGCAAG 861
 Db 718 GAGTAAG 777
 QY 862 GACCAAG 921

Db 778 CACCAAG 837
 QY 922 CAG 981
 Db 838 CAG 897
 QY 982 CAG 1041
 Db 898 TGAAG 957
 QY 1042 GCGTGGCTCCCTCCGTTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101
 Db 958 ACCGAGATCCCTTCGTTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017
 QY 1102 TCGATCTCGCTCCAG 1161
 Db 1018 TCGCTCTCGCTCCAG 1077
 QY 1162 GAGCCGCGCTCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 1221
 Db 1078 GAGCCGAG 1137
 QY 1222 ACCGCTCCGAG 1281
 Db 1138 ACCGCTCCGAG 1197
 QY 1282 CAG 1341
 Db 1198 CAG 1257
 QY 1342 CCGCATCCCTCGAG 1401
 Db 1258 CACCAAG 1317
 QY 1402 CCGATCCCTCGAG 1461
 Db 1318 CCGATCCCTCGAG 1377
 QY 1462 CCGAG 1518
 Db 1378 CCGAG 1437
 QY 1519 GCGAG 1578
 Db 1438 GAG 1497
 QY 1579 TGAAG 1638
 Db 1498 GGAAG 1557
 QY 1639 CCGCTCCAG 1698
 Db 1558 CC-----CAG 1603
 QY 1699 TATTATTCTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT 1758
 Db 1604 TATTATTCTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT 1663
 QY 1759 CACTAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1818
 Db 1664 CACTAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1723
 QY 1819 T 1819
 Db 1724 T 1724

RESULT 8
 AF358134 1671 bp mRNA linear ROD 29-MAY-2001
 LOCUS AF358134
 DEFINITION Mus musculus naked cuticle-1 (Nkd1) mRNA, complete cds.
 ACCESSION AF358134

VERSION AF358134.1 GI:14211711
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
1. 1671
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6xCBA"
/db_xref="taxon:10090"
/clone="L84"
/sex="female"
/issue_type="lung"
/dev_stage="6-8 weeks"
1. 1671
/gene="Nkd1"
80. 1495
/note="contains EFX domain, similar to Drosophila melanogaster naked cuticle (nkd)"
/product="naked cuticle-1"
/protein_id="AAK57483.1"
/db_xref="GI:14211712"
/translation="MGKLHSPRAVCCRRSPEDSDPAVSAMARKIEWIGRORCP
GSVSPQLRVLAVGRTRELVGDVSRREALGEDEDPLEVALPREKIDLSGDE
KRMRLESEPOASKKQKPEBLQCDVVEEDSRQEMFTLYDPNNKVRBEDTSL
HTIYEVVDSVNHPSSTKTLRVLTVAIPGDSQSKRSVLFNHTDLSRPA
ELRGMEKKORAPLRFQDSHLEODCYHNCVDENIERNNYLDLAGLENTSPADP
PSVAOKSELPRISNPIRPSRSHPEPAHIIHRPQGVPSFHLDPAPKASFLOR
LRGTQDSKIFVSPKAOQKMGHGAARARSRPVPYTHYTSPEAHNTSPALP
TLAVLGHKKHKKHAKESQASCRGLQGPLAAGSTVWGREDVRELPAVMVYESQGVAV
QRHHNNHHNNHHYHNP"OP"

ORIGIN
Query Match 64.3%; Score 1194.8; DB 10; Length 1671;
Best Local Similarity 84.4%; Pred. No. 2e-206;
Matches 1398; Conservative 0; Mismatches 237; Indels 21; Gaps 4;
QY 167 CAGGCCGCGGCGGCGGCGATGAGGAGCGCTCCGCGCGCGGCGAGCCCGAGCATG 226
DB 27 CGGGGTGGGCGGCGGCGAGCGATGAGCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 82
QY 227 GGGAACTTCACTCAAGCCGCGCGCGCTGTGCAAGCCGAGGAGAGAGCCCGGAAAGTGAC 286
DB 83 GGGAACTTCACTCAAGCCGCGCGCGCTGTGCAAGCCGAGGAGAGAGCGCGGAAAGTGAC 142
QY 287 AGCTTCCCGTGAACCGCTGCTGGGCTCGGAAAGGAGCATCGAGAGTGATGGGAGACAG 346
DB 143 AGCTTCCCGTGAACCGCTGCTGGGCTCGGAAAGGAGCATCGAGAGTGATGGGAGACAG 202
QY 347 CGCTGCCGCGGCGGCTGTGCGGAGCCCGAGCAGCTGCGGCGGCGAGCCCGAGCATGAGCGCA 406
DB 203 CGCTGTCCAGCAGCGCTTCAAGAACCCGCTGAGCTGAGATTGGGAGGCGAGCTGTGGTCCA 262
QY 407 AGCAGCCGAGGAGCTGTGGGCGAGCGTGTGAGAGCAGCTCAGCAGGAGAGAGAGAGAC 466
DB 263 GGCAGCTGGGAACTCGTGGGTGAGACATTCTAGAGAGGCTCTCGGTGAGAGAGAGAGAGAC 322

QY 467 GACTTTCGCTGGAAGTGGCCCTGCTCTGAGAGACTGACGAGGCTGGCAGCGAGAT 526
DB 323 GACTTCCCCCTAAGAGTGGCCCTGCGCCCTGAGAGAGATGACAGCGCTAGGTAGAGAT 382
QY 527 GAGAGAGAGTGAAGAGTGAAGCCCTGCTCCAGGCTTCAAGAGAGAGTGAAGTTT 586
DB 383 GAGAGAGAGTGAAGAGTGAAGCCCTGCTCCAGGCTTCAAGAGAGAGTGAAGTTT 442
QY 587 GAAGAGCTCAGTGGCCAGCGTGTCCATGAGAGAGAGAGAGAGAGTGAAGTGAAGC 646
DB 443 GAAAGCTACAGTGTGTCTGTGTGAGAGAGAGAGAGAGAGAGTGAAGTGAAGTGAAGT 502
QY 647 CTGATGACTTTTGAACAACAGGCAAGGTCAACCCGAGAGAGACATCAACCTTGTGAC 706
DB 503 CTATATGACTTGAACAACAGGCAAGGTCAACCCGAGAGAGACATCAACCTTGTGAC 562
QY 707 ACCATCTATGAGTGTGTGAGCTGCTGTGCAACCACTCCCAATCAGAGAGAGTGTG 766
DB 563 ACCATCTATGAGTGTGTGAGCTGCTGTGCAACCACTCCCAATCAGAGAGAGTGTG 622
QY 767 CGGTTAAAGTCAACCGTGGCCCGGATGTGAGAGAGAGAGAGAGAGTGTGTGAT 826
DB 623 CGGTTAAAGTCAACCGTGGCCCGGATGTGAGAGAGAGAGAGAGAGTGTGTGAT 682
QY 827 CAGGCTGACCTGGAGAGCGGAGGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 886
DB 683 CATACGATTTGAG 742
QY 887 AGCTGGAG 946
DB 743 GCTGGAG 802
QY 947 TCTGGCTGCTTCAACCTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1006
DB 803 CAGAGCTGCTTCAACCTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 862
QY 1007 CTGCGCGGAGTGAAGAACTCAAGTCCCAATTTGGGCTGCTGCTGCTGCTGCTGCTG 1066
DB 863 CTGCGCGGAGTGAAGAACTCAAGTCCCAATTTGGGCTGCTGCTGCTGCTGCTGCTG 922
QY 1067 AAGTCAGAACTGCG 1126
DB 923 AAGTCAGAACTGCG 982
QY 1127 GGCATCACTTCCCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
DB 983 GCTGCCCACTCCCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1042
QY 1187 GACACCCCAATGCGCAAGGTCTGAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
DB 1043 GACACCCCAATGCGCAAGGTCTGAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102
QY 1247 AGCAAGCACTTGTGAGTCCCGCAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306
DB 1103 AGCAAGCACTTGTGAGTCCCGCAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162
QY 1307 GCGAG 1366
DB 1163 GCGAG 1222
QY 1367 GCCCACTGAGTCCAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1426
DB 1223 GCCCACTGAGTCCAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1282
QY 1427 CACAAGCACTGAGTCCAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486
DB 1283 CACAAGCACTGAGTCCAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1342
QY 1487 TCAAGTGG---CCCTGTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1543
DB 1343 GCAAGAGAGTCCAGCGCTGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1402

QY	1544	TATGAGACCCGAGCGCCGAGCCGCTCCAGAGATGAGACACACACACACATGAACAT	1603
Db	1403	TACGAGACCCAGGCTGGGACAGCGCTCCAGAGACACAGAACATCACACACACAGAACAT	1462
QY	1604	CACCACCATTAACCAACACTTCTACAGACATAGAGCCCTCCGAGGAGCCCAACCTGCC	1663
Db	1463	CACCACCATTAACCAACACTTCTATACAGCCCTTAGAGCC-----CAGCAGGCTGCC	1511
QY	1664	ATATGAAGAGACCCCAACCCCGACACACACAGGACATTAATTCTAATTATTTGTTATTT	1723
Db	1512	ACGGGAAGAGACCCAGCCC---ACACCCCTAAGGACATTAATTATCTAATTATTTGTTATTT	1568
QY	1724	ATGATGATTTATTTGTTAATTAATTAATTTATTTGTTACTTACCTAATATTTTGCTGCTACATG	1783
Db	1569	ATGCAATTAATTTGTTAATTAATTAATTTATTTGTTACTTACCTAATATTTTAGCCAGCTTATG	1628
QY	1784	TAGAAGATCTATGTGGAACACAGAACTAACTTTTAT	1819
Db	1629	TAGAAGACATCATGGAACACAGAACTAACTTTTAT	1664
RESULT 9	AF343352	1416 bp	linear
LOCUS	Mus musculus Nkd mRNA, complete cds.		ROD 29-MAR-2001
DEFINITION	AF343352		
ACCESSION	AF343352.1	GI:13487304	
VERSION			
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1416)		
AUTHORS	Yan, D., Wallingford, J.B., Sun, T.-O., Nelson, A.M., Sakanaka, C., Reinhard, C., Harland, R.M., Fanti, W.J. and Williams, L.T.		
TITLE	Cell autonomous regulation of multiple Dishevelled-dependent pathways by mammalian Nkd		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 3802-3807 (2001)		
MEDLINE	21173639		
REFERENCE	2 (bases 1 to 1416)		
AUTHORS	Yan, D., Wallingford, J.B., Sun, T.-O., Nelson, A.M., Sakanaka, C., Reinhard, C., Harland, R.M., Fanti, W.J. and Williams, L.T.		
TITLE	Direct Subdomain		
JOURNAL	Submitted (26-JAN-2001) Research, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA		
FEATURES	location/Qualifiers		
source	1. 1416		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	1. 1416		
	/notes="mMk: similar to Drosophila naked cuticle"		
	/codon_start=1		
	/product="Nkd"		
	/protein_id="AAK27485.1"		
	/db_xref="GI:13487305"		
	/translation="MGKLSHFAAVNCRRRESPEDSFAVAAAMARKIEIMIGRQCPGSVAGPROLRAGTGRVETLVDGTRREALGEEDDPRLVLPKEDISLGGGEKMRRLSEPGQASKKQIKFELDQVDEDSROEMFTLYDPNNKKTREDTSLHTIIEVDSSVNSPRTSSKTRVLTVAFDQSSKRSVLFRNHDLDSTRPADTKPKEELRHEKKQRAVRLPQGDSDHLEQDCVNHCDENIERNNHLDLAGIENTVSGPQSSPSVAQKSELPRPISNTPRSRHEDEAAHPIRRPGVDPGSFILDTPEKASLQORLRGQDQSKTVRSRPKAGKMGWGSHGARGRSKRPVPTTHVPSANHLATSPALPRLAFLGKHKHRAKESQASCRGLDQFLAAGGSIVMREQVRELPAVVVVSQAQAVVQRNHNHNHNHNHNHNHNYQR"		
ORIGIN			
Query Match	58.0%;	Score 1078.4;	DB 10; Length 1416;
Best Local Similarity	85.6%;	Pred. No. 2.6e-185;	
Matches 1212;	Conservative 0;	Mismatches 201;	Indels 3; Gaps 1;
QY	224	ATGGGGAATTCACCTCCAGCCGCGCGCGCTGTGCAAGCCGACGAGAGACCCCGAAAGT	283

Db	1	ATGGGAAACTTCACTCGAAGCCGGGCGCCCTGTGCAAGCCGAGGAGAGCCCGGAAGT	60
QY	284	GACAGCTTCCGCGTGAAGCGCTGCTGGGCTCGAAGGGCATCGAGAGTGAATCGGAGA	343
Db	61	GACAGCTTGTGCTGAAGGCTGTGGCAAGAAAGGCATCGAGAGTGAATCGGAGG	120
QY	344	CAGGCGTCCCGGCGGTGTCTCGGGAATCCCGACAGCTGCGGTGGCGGCACTAAGC	403
Db	121	CAGGCGTGTCCAGGCAAGGCTCAAGGAATCCCGTAGCTGAATTTGGCAGGGACTGTGTGT	180
QY	404	CGAAGCAATCCGGAGCTGTGGGAGACGTGTTGAGAGACAGCTCAGGAGAGAAAGGAG	463
Db	181	CGAGGACTCGGGAACTGTGGGTGACATTTGAGAGGCTTCTCGGTGAGAGAGAGAG	240
QY	464	GACGACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAAAGTGAACGGCTGGGACGCGA	523
Db	241	GACGACTTCCCCATGAAGGTGGCCCTGCGCTGAGAAATCGACAGCTGATGATGGA	300
QY	524	GATGAGAAAGATGAGAGAGTGAAGCAATCCCTGCGGCTCCAGAACAGTGAAG	583
Db	301	GATGAGAAAGAAATGAGAGAGACTAGAGAACTTGGCCAGGCTCCAGAAACAGTCAAG	360
QY	584	TTTGAAGAGCTCCAGTGGACGTGTCCATGGAAGAGGACAGCGGAGAGATGGACCTTC	643
Db	361	TTTGAAGAGCTACAGTGTGATGTCTCTGTGAGAGAGAGCAGCGGAGAGATGGACCTTC	420
QY	644	ACCCTGTATGACTTTTGAACAACAAGGAGTCAACCGAGAGAGACATCACGACTTGCTG	703
Db	421	ACTGTATATGACTTGCACACAAATGAGCAAAATGACCGGTAGAGACTTATCAAGCTTG	480
QY	704	CACACCATCTATGAGGTGTGAGCTCCTCTGTCAACAATCCCTCAATCCAGACAGTG	763
Db	481	CATACCAATCTATGAGGTGTGAGCTCCTCTGTGAAACAATCCCACTCAAGCAAGACA	540
QY	764	CTGGGGGTAAGCTCACCGTGGCCCCCGATGGCCAGAGACAGAGAGAGGTCTTGTTC	823
Db	541	CTGGGGGTAAGCTCACCGTGGCTTCCTGACGGAGCCAGATTAAGAGAGGTCTTGTTC	600
QY	824	AATCAGGCTGACTGACAGAGCGCAAGGCCCGAGACAGAGCAAGACCATGAGGACCTG	883
Db	601	AACCATACCGACTGTGCAGAGACCAAGGCCCGAGACAGACCAAAATCCGCTGAGAGAGTG	660
QY	884	CGAGAGTGGAGAGAAAGCAGCGAGCCCGCTCAGGTTCCAGGGTGAACGCCGCTTGAG	943
Db	661	CGTGGCTGGAGAGAAAGCAGCGAGCCCACTCAGGTTCCAGGGTGAACGCCCACTTGAG	720
QY	944	CAGTCTGGCTGTACACCAATTTGGGTATGATGAAACATCGAGAGAGAAATCACTACTTA	1003
Db	721	CAGCCAACTGTCTACACATTTGGGTATGATGAAATGAGAGAGAAATCACTACTTA	780
QY	1004	GATCTCGCCGGATAGAAAATCTACAGCGCCCAATTTTGGGCTGTGCTCCCTTCGTTGCC	1063
Db	781	GACTTGGCGGGATATGAAATCTACAGCTGTCAAGTTTGAACCGGATATCCCTTCGTTGCC	840
QY	1064	CAGAAGTCAAGACTGCCGCCCGCAGCACTCCAAATCCAATCGATGTGCTTCATGAGCGG	1123
Db	841	CAGAAGTCAAGACTGCCGCCCGCAGCACTCCAAATCCAATCGATGTGCTTCATGAGCGG	900
QY	1124	GAAACCATCCATATCCCAATCCGAATAGCCCAAGCGGTGAACCCGGGCTCTCTTCACTTC	1183
Db	901	GAAACCATCCCAATATCCCAATAGCCCAAGCGGTGAACCCCAAGCTCTCTTCACTTC	960
QY	1184	CTTACACATCCCAATATCCCAAGGTCTCAGAGCTCAGAGCAAGGTCTCCGAGGACCCAGGAC	1243
Db	961	CTTACACATCCCAATATTCGCAAGGACTCAGAGCTCAGAGCAAGGTCTCCGAGGACTCAGGAT	1020
QY	1244	GGAAGCAAGCACTTTGTGAGGTCCCGCAAGGCCCAAGGCAAGAGTGTGTGTGTGGCCAC	1303
Db	1021	GGAAGCAAGCACTTTGTGAGGTCCCGCAAGGCCCAAGGCAAGAGTGTGTGTGTGGCCAC	1080
QY	1304	GTGGCCAGAGGGGCAAGAAACAAAGCCCCCTCTGTGGACATCCGCTCTCTGTGTGTGTGT	1363

Dh 1081 GGGGCCAGAGTGCAGAAAGCAAGCCTCACTGTGTACCAACCAACCATATCTGTCTCCCC 1140
Qy 1364 TCGGCCCACTGCTGTCAGAGCCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1423
Db 1141 TCTGCCCATCTGGGCAACAGCCGAGCCCTTCTCTCCCAACCTGTGCACTCTCTCTCTCTCTCT 1200
Qy 1424 AAGCAAGAGCAGAGCCAGAGAGCAGAGAGGCTGCGGGGCTGAGGAGCACTG 1483
Db 1201 AAGCAAGAGATGAG 1260
Qy 1484 GCCTCAGAGTGG---CCCTGTCTGTGGGGCGAGAGCAGCTGCGAGAGCTGCGCCCTTGGTG 1540
Db 1261 GCTGAGAGAGAGCTCACTGTCTATGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1541 GTGTATAGAGAGCCAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
Db 1321 GTGTACAGAGAGCCAGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1601 CATCACCACTTATACCACTTCTACAGAGATAG 1636
Db 1381 CATCACCACTTATACCACTTCTATACAGCCTTAG 1416

RESULT 10

AX172351 1401 bp DNA linear PAT 03-JUL-2001
LOCUS AX172351
DEFINITION Sequence 1 from Patent WO0144279.
ACCESSION AX172351
VERSION AX172351.1 GI:14597503
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Van, D. and Williams, L. T.
TITLE Mammalian dishevelled-associated proteins
JOURNAL Patent: WO 0144279-A 1 21-JUN-2001;
Chiron Corporation (US)
FEATURES
SOURCE 1. 1401
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

ORIGIN

Query Match 57.4%; Score 1066.6; DB 6; Length 1401;
Best Local Similarity 85.6%; Pred. No. 3.6e-183;
Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

Qy 224 ATGGGGAAATCTTCACTCCAGCCGCGCGCTGTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGT 283
Db 1 ATGGGGAAATCTTCACTCCAGCCGCGCGCTGTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGT 60
Qy 284 GACAGCTTCCGCGAG 343
Db 61 GACAGCTTCCGCGAG 120
Qy 344 CAGCGCTGCGCGAG 403
Db 121 CAGCGCTGCTCAG 180
Qy 404 CGAAGCAACCGGAG 463
Db 181 CGAAGCAACCGGAG 240
Qy 464 GACGACATTTTGGCTGAG 523
Db 241 GACGACATTTTGGCTGAG 300
Qy 524 GATGAGAGAGAGATGAG 583
Db 301 GATGAGAGAGAGATGAG 360

Qy 584 TTTGAAGAGCTCGAGTGCAG 643
Db 361 TTTGAAGAGCTACAGTGTGATGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy 644 ACCCTGATGACTTTTGAACAACAAG 703
Db 421 ACTCTATATGACTTTGAG 480
Qy 704 CACACATCTATGAG 763
Db 481 CATACCATCTATGAG 540
Qy 764 CTGCGGGTAAAGTCAACCGTGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823
Db 541 CTGCGGGTAAAGTCAACCGTGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 824 AATCAGGCTGACCTGAG 883
Db 601 AATCAGGCTGACCTGAG 660
Qy 884 CGAGAGCTGGAG 943
Db 661 CTGAGCTGGAG 720
Qy 944 CAGTCTGCTGCTTACCAACCACTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1003
Db 721 CAGCAGAGAGCTTACCAACCACTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 1004 GATCTGCGCGGAGATGAG 1063
Db 781 GACTGCGCGGAGATGAG 840
Qy 1064 CAGAGAGTGAAGTGTGAG 1123
Db 841 CAGAGAGTGAAGTGTGAG 900
Qy 1124 GAAGCCATCCAGATCCAG 1183
Db 901 GAAGCCATCCAGATCCAG 960
Qy 1184 CTTGACACCCCAATTCGCAAGAGTCTCAG 1243
Db 961 CTTGACACCCCAATTCGCAAGAGTCTCAG 1020
Qy 1244 GAGAGCAAGCACTTTGTGAGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1303
Db 1021 GAGAGCAAGCACTTTGTGAGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1304 GTGGCCAG 1363
Db 1081 GGGGCGCAAGAGTGTGAAG 1140
Qy 1364 TCGGCCCACTGCTGTCAGAGCCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1423
Db 1141 TCTGCCCATCTGGGCAACAGCCGAGCCCTTCTCTCCCAACCTGTGCACTCTCTCTCTCTCTCTCT 1200
Qy 1424 AAGCAAGAGCAGAGCCAG 1483
Db 1201 AAGCAAGAGATGAG 1260
Qy 1484 GCCTCAGAGTGG---CCCTGTCTGTGGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1540
Db 1261 GCTGAGAGAGAGCTCACTGTCTATGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1541 GTGTATAGAGAGCCAGGCGCGGAG 1600
Db 1321 GTGTACAGAGAGCCAGGCTGGGAG 1380
Qy 1601 CATCACCACTTATACCACTTCTACAGAGATAG 1636
Db 1381 CATCACCACTTATACCACTTCTATACAGCCTTAG 1401

[illegible]

OY	1136	ATCCACACCGAAAGCCCCCAAGGGGTGACCCGGGCTCTTCCACTTCCTTGAACACCCA	11935
Db	736	ATCCACACCGAAAGCCCCCAAGGGGTGACCCGGGCTCTTCCACTTCCTTGAACACCCA	795
OY	1196	ATCCCAAGGCTCAGAGCTTCAGCAACCGCTCCGGGGACCCACGAGACGGAGCAAGCAC	12555
Db	796	ATCCCAAGGCTCAGAGCTTCAGCAACCGCTCCGGGGACCCACGAGACGGAGCAAGCAC	855
OY	1256	TTGTGAGTGTCCCCCAAGGCCCAAGGCAAGAGTGTGAGTGTGGGCCACGTGGCCAGAGGG	13155
Db	856	TTGTGAGTGTCCCCCAAGGCCCAAGGCAAGAGTGTG-----	891
OY	1316	GCAAGAAACAAGCCCCCTCTGGGACCCGGCATTCCCTGGGTCCCTCCGCCACCTG	13755
Db	892	-----	891
OY	1376	GCTGCAGACCCCGGCTCTCTCCCTCCCTAGCCCCCTCGGGACAAAGACAAACAC	14355
Db	892	-----GCCCCCTCGGGACAAAGACAAACAC	921
OY	1436	CGACCCAAAGAGAGCCACGAGGCTGCGGGGCTTCAGAGCACCATGCGCTCAGGTGGC	14955
Db	922	CGACCCAAAGAGAGCCACGAGGCTGCGGGGCTTCAGAGCACCATGCGCTCAGGTGGC	981
OY	1496	CCGTCTCCGGGGGGGGAGACCTGCGGGAGCTGCGCGCTTGAGTGATGAGAGCGAG	15555
Db	982	CCGTCTCTGGGGGGGGAGACCTGCGGGAGCTGCGCGCTTGAGTGATGAGAGCGAG	10411
OY	1556	GCCCGGACAGCCG 1567	
Db	1042	GCCGGGACAGCCG 1053	
RESULT 12			
LOCUS	HSA303140	16319 bp	DNA linear PRI 03-APR-2004
DEFINITION	Homo sapiens NOD2 gene for LRR-containing protein, exon 1-11.		
ACCESSION	AJ303140		
VERSION	AJ303140.1 GI:14488148		
KEYWORDS	LRR-containing protein; NOD2 gene.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Hugot,J.P., Chamaillard,M., Zouali,H., Lesage,S., Cezard,J.P.,		
	Belache,J., Almer,S., Tysk,C., O'Morain,C.A., Gassull,M.,		
TITLE	Binder,V., Finkel,Y., Cortot,A., Modigliani,R., Laurent-Puig,P.,		
	Gower-Rousseau,C., Macry,J., Colombel,J.F., Sahbatou,M. and		
JOURNAL	Thomas,G.		
	Association of NOD2 leucine-rich repeat variants with		
MEDLINE	susceptibility to Crohn's disease		
	Nature 411 (6837), 599-603 (2001)		
PUBMED	21279172		
	11385576		
REFERENCE	2 (bases 1 to 16319)		
	Zouali,H.		
TITLE	Direct Submission		
	Submitted (02-JAN-2001) Zouali H., Genetics, Fondation Jean		
JOURNAL	Dausset-CEPH, 27 RUE Juliette Dodu, 75010 Paris France, FRANCE		
	Location/Qualifiers		
FEATURES	1..16319		
	/organism="Homo sapiens"		
SOURCE	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
gene	/chromosome="16"		
	/map="16q12-13"		
mrna	/clone_1b="hb87b10"		
	/clone_1b="CEPH Human BAC library"		
	98302..131994		
	/gene="NOD2"		
	join(98302)..98906,106807..106912,109510..111325,		

cds
 join(984448..98906,106807..106912,109510..111325,
 115517..115600,115821..115904,118855..118938,
 121552..121635,122230..122313,124418..124501,
 128745..128828,130674..131994)
 /gene="NOD2"
 98302..98906
 /gene="NOD2"
 /number=1
 /function="putative NF-kB activator"
 /note="protein of the CARD family"
 /codon_start=1
 /product="LRR-containing protein"
 /protein_id="CAC42117.1"
 /db_xref="GI:1448149"
 /db_xref="GOA:O9HC29"
 /db_xref="Swiss-Prot:O9HC29"
 /translation="MCSDEAFQAKRSQLVELLVSGSLGEPESVLDMLSEVSEWEDY
 EEFHLLGGLSLARRLLDPTWNNKGTAMCOKLIAAQAPQADSOSPKLHGCDWDSHLH
 PARDQSHRPATVRRHSHVEMLDLMEARGVSQECDELRLPIFPSPRARVLLDL
 ATVKANGAAALPLGHVQELPYPLALPIEAATCCYTMKALRTTVSAQSRFLSTYGATT
 LCLEIDITYENVEVWADVGMAGSPQKSPATIGLEELFSTPEHLNDDADTVLVSEAGS
 GSTFLEHRLHLLMAAGDOFOEFLFVPPSCRLQCMAPTSVRTLFEHCWPDVQGE
 DIFOLLDRPDEVLLTFDGFDFEKKFPFDRRHGSPDPTSVOTLLFNLQSNLKNAA
 RLVHSPCHPAVSAPFLRKYIRTEFPNLKGSBEOGIELYLRKHHEGVDRLRLIQETS
 ALHGICRLPVPSPWSPKCHOELLQEGGSPKTTDMYLLIIOHFLNATPPDSASQGL
 GSSLRGRLPTLHLGRALNKGCMCCVFSNQLQAQVSPDISIGFLVRAGVYP
 GSGTAPLEPHITIQCEPFAFYLASADVPALRLRLFVCGSPGNSPMARLLPTMCIA
 SGTQSSVALVLLQCAEPHNLQTTAFLAGLSREHMGILACQTSERKALRRQCAW
 CLARSLRKFIHSLIPPAAPGEAKSVAMGFTLIRSLTYEMOBERLAKAARGLVGHIL
 KLTGVSVEGTCEALAFVLOHRRPVALLDYNYSVDIGVOLLPCLGVCALVLRDNYI
 NISDGICTCEALCAVLOKTLFNNKLLDGCANSMARKLLARONFLAIRGNNTYI
 TPAAGQVLABGKRGTSIQFLGFMKRGVDEGAQALALADHQSILKRLSIVGNNTSS
 VAAQALATLAKNVMLEELCLEENHLQDEGVCSLAEGKKNSSLKILKLSNNCTITLGG
 AEAALLQALERNNTIILEVLRGNTFSLIEVDKLCGRDRLLL"
 98907..106806
 /gene="NOD2"
 /number=1
 106807..106912
 /gene="NOD2"
 /number=2
 106913..109509
 /gene="NOD2"
 /number=2
 109510..111325
 /gene="NOD2"
 /number=3
 111326..115516
 /gene="NOD2"
 /number=3
 115517..115600
 /gene="NOD2"
 /number=4
 115601..115820
 /gene="NOD2"
 /number=4
 115821..115904
 /gene="NOD2"
 /number=5
 115905..118854
 /gene="NOD2"
 /number=5
 118855..118938
 /gene="NOD2"
 /number=6
 118939..121551
 /gene="NOD2"
 /number=6
 121552..121635
 /gene="NOD2"
 /number=6
 121635..122230
 /gene="NOD2"
 /number=6
 122313..124418
 /gene="NOD2"
 /number=6
 124418..124501
 /gene="NOD2"
 /number=6
 124501..128745
 /gene="NOD2"
 /number=6
 128745..128828
 /gene="NOD2"
 /number=6
 128828..130674
 /gene="NOD2"
 /number=6
 130674..131994
 /gene="NOD2"
 /number=6
 131994..138745
 /gene="NOD2"
 /number=6
 138745..148745
 /gene="NOD2"
 /number=6
 148745..158745
 /gene="NOD2"
 /number=6
 158745..168745
 /gene="NOD2"
 /number=6
 168745..178745
 /gene="NOD2"
 /number=6
 178745..188745
 /gene="NOD2"
 /number=6
 188745..198745
 /gene="NOD2"
 /number=6
 198745..208745
 /gene="NOD2"
 /number=6
 208745..218745
 /gene="NOD2"
 /number=6
 218745..228745
 /gene="NOD2"
 /number=6
 228745..238745
 /gene="NOD2"
 /number=6
 238745..248745
 /gene="NOD2"
 /number=6
 248745..258745
 /gene="NOD2"
 /number=6
 258745..268745
 /gene="NOD2"
 /number=6
 268745..278745
 /gene="NOD2"
 /number=6
 278745..288745
 /gene="NOD2"
 /number=6
 288745..298745
 /gene="NOD2"
 /number=6
 298745..308745
 /gene="NOD2"
 /number=6
 308745..318745
 /gene="NOD2"
 /number=6
 318745..328745
 /gene="NOD2"
 /number=6
 328745..338745
 /gene="NOD2"
 /number=6
 338745..348745
 /gene="NOD2"
 /number=6
 348745..358745
 /gene="NOD2"
 /number=6
 358745..368745
 /gene="NOD2"
 /number=6
 368745..378745
 /gene="NOD2"
 /number=6
 378745..388745
 /gene="NOD2"
 /number=6
 388745..398745
 /gene="NOD2"
 /number=6
 398745..408745
 /gene="NOD2"
 /number=6
 408745..418745
 /gene="NOD2"
 /number=6
 418745..428745
 /gene="NOD2"
 /number=6
 428745..438745
 /gene="NOD2"
 /number=6
 438745..448745
 /gene="NOD2"
 /number=6
 448745..458745
 /gene="NOD2"
 /number=6
 458745..468745
 /gene="NOD2"
 /number=6
 468745..478745
 /gene="NOD2"
 /number=6
 478745..488745
 /gene="NOD2"
 /number=6
 488745..498745
 /gene="NOD2"
 /number=6
 498745..508745
 /gene="NOD2"
 /number=6
 508745..518745
 /gene="NOD2"
 /number=6
 518745..528745
 /gene="NOD2"
 /number=6
 528745..538745
 /gene="NOD2"
 /number=6
 538745..548745
 /gene="NOD2"
 /number=6
 548745..558745
 /gene="NOD2"
 /number=6
 558745..568745
 /gene="NOD2"
 /number=6
 568745..578745
 /gene="NOD2"
 /number=6
 578745..588745
 /gene="NOD2"
 /number=6
 588745..598745
 /gene="NOD2"
 /number=6
 598745..608745
 /gene="NOD2"
 /number=6
 608745..618745
 /gene="NOD2"
 /number=6
 618745..628745
 /gene="NOD2"
 /number=6
 628745..638745
 /gene="NOD2"
 /number=6
 638745..648745
 /gene="NOD2"
 /number=6
 648745..658745
 /gene="NOD2"
 /number=6
 658745..668745
 /gene="NOD2"
 /number=6
 668745..678745
 /gene="NOD2"
 /number=6
 678745..688745
 /gene="NOD2"
 /number=6
 688745..698745
 /gene="NOD2"
 /number=6
 698745..708745
 /gene="NOD2"
 /number=6
 708745..718745
 /gene="NOD2"
 /number=6
 718745..728745
 /gene="NOD2"
 /number=6
 728745..738745
 /gene="NOD2"
 /number=6
 738745..748745
 /gene="NOD2"
 /number=6
 748745..758745
 /gene="NOD2"
 /number=6
 758745..768745
 /gene="NOD2"
 /number=6
 768745..778745
 /gene="NOD2"
 /number=6
 778745..788745
 /gene

Query Match	Beat Local Similarity	42.1%; Score 783; DB 9; Length 163319;	Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
intron	/number=7 121636..122229 /gene="NOD2"		
exon	/number=7 122230..122313 /gene="NOD2"		
intron	/number=8 122314..124417 /gene="NOD2"		
exon	/number=8 124418..124501 /gene="NOD2"		
intron	/number=9 124502..128744 /gene="NOD2"		
exon	/number=9 128745..128828 /gene="NOD2"		
intron	/number=10 128829..130673 /gene="NOD2"		
exon	/number=10 130674..131994 /gene="NOD2"		
ORIGIN	/number=11		
Query Match	42.1%; Score 783; DB 9; Length 163319;		
Beat Local Similarity	100.0%; Pred. No. 1.2e-131;		
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
1046	GGCTCCCTTCCTCGTGGCCCGAGAGTCAAACTGCCCCCGCACCTTCAATCCCACTCCGA	1105	
31966	GGCTCCCTTCCTCGTGGCCCGAGAGTCAAACTGCCCCCGCACCTTCAATCCCACTCCGA	3202	
1106	TCTGGCTCCCATGAGCCGGGAAGCCATCCATCCCAACCGAAGCCCAAGCGGTGAGAC	1165	
32026	TCTGGCTCCCATGAGCCGGGAAGCCATCCATCCCAACCGAAGCCCAAGCGGTGAGAC	3208	
1166	CCGGCTCTCTTCACCTTCTCTTGAACACCCCAATCGCAAGGTCTTCAAGCTTCAAGCAACG	1225	
32086	CCGGCTCTCTTCACCTTCTCTTGAACACCCCAATCGCAAGGTCTTCAAGCTTCAAGCAACG	3214	
1226	CTCCGGGGCACCCAGACGGGAGCACTTTGTGAGTTCGCCCAAGGCCCAAGGGCAAG	1285	
32146	CTCCGGGGCACCCAGACGGGAGCACTTTGTGAGTTCGCCCAAGGCCCAAGGGCAAG	3220	
1286	AGTGTGGGTGTGGGCAACGTGGCCAGAGGGGCAAGAAACAAGCCCTTGTGGACCCGCC	1345	
32206	AGTGTGGGTGTGGGCAACGTGGCCAGAGGGGCAAGAAACAAGCCCTTGTGGACCCGCC	3226	
1346	ATCCCTGAGGTATCCCCCTCCGCCACCTGGCTGCAGACCAGGCCCTTCTCCCTCTTA	1405	
32266	ATCCCTGAGGTATCCCCCTCCGCCACCTGGCTGCAGACCAGGCCCTTCTCCCTCTTA	3232	
1406	GCCCCCTCTGGGGCAAGAAAGCAACAAGCCAGACCCAGAGAGACCCAGAGGGTGCAG	1465	
32326	GCCCCCTCTGGGGCAAGAAAGCAACAAGCCAGACCCAGAGAGACCCAGAGGGTGCAG	3238	
1466	GACCTGACAGCACTAGGCTCAAGGTGGCTTGTCTTGGGGGGGAGAGCACTTGGGGAG	1525	
32386	GACCTGACAGCACTAGGCTTCAAGGTGGCTTGTCTTGGGGGGGAGAGCACTTGGGGAG	3244	
1526	CTGCCCGCTTGTGGGTATGAGAGCGAGCGGGAGCGGTGTCAGAGCATGAGCAC	1585	
32446	CTGCCCGCTTGTGGGTATGAGAGCGAGCGGGAGCGGTGTCAGAGCATGAGCAC	3250	
1586	CACCAACAACATGAACATCAACACATTTACCACTTACACAGATAGAGCCCTCC	1645	
32506	CACCAACAACATGAACATCAACACATTTACCACTTACACAGATAGAGCCCTCC	3256	
1646	CGAGGGCCCACTTGCATATGAAGACCCACCCCGACACACAAGGATTTATTT	1705	

Db 3256 CCAAGGCCCAACCTGCGATATGAGAGACCCCAACCCGACACACAGGATTATAT 32625
Qy 1706 CTAATTAATTGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1765
Db 32626 CTAATTAATTGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 32685
Qy 1766 ATTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 1825
Db 32686 ATTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 32745
Qy 1826 GTT 1828
Db 32746 GTT 32748

RESULT 13
AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
LOCUS Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
DEFINITION AC007608
AC007608.7 GI:29124038
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169773)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,D., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., McBride,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
REFERENCE 3 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 5 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Mar 19, 2003 this sequence version replaced gi:18129388.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1.169773
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-401P9"

FEATURES
source
ORIGIN

Query Match 42.1%; Score 783; DB 9; Length 169773;
Best Local Similarity 100.0%; Pred. No. 1,2e-131;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1046 GGCCTCCCTCCGAGGCCCAAGATGCAAGTGGCCCCCGGACCTCCATCCACTGCA 1105
Db 78524 GGCCTCCCTCCGAGGCCCAAGATGCAAGTGGCCCCCGGACCTCCATCCACTGCA 78465
Qy 1106 TCTGCTCCCATGAGCCGGAAGCATCCATCCACACCGAAAGCCCAAGGCTGGAC 1165
Db 78464 TCTGCTCCCATGAGCCGGAAGCATCCATCCACACCGAAAGCCCAAGGCTGGAC 78405
Qy 1166 CCGGCTCTTCCATCTTCTTGAACCCCAATGGCCAAAGTCTTCAGACTTCAGAAACG 1225
Db 78404 CCGGCTCTTCCATCTTCTTGAACCCCAATGGCCAAAGTCTTCAGACTTCAGAAACG 78345
Qy 1226 CTCGGGGGACCCAGAGCGGGAGCAAGCACTTTGTGAGTCCCGCAAGGCCAGGGGACG 1285
Db 78344 CTCGGGGGACCCAGAGCGGGAGCAAGCACTTTGTGAGTCCCGCAAGGCCAGGGGACG 78285
Qy 1286 AGTGTGGTGTGGGACCACTGTGGCCAGAGGGGCAAGAAACAGGCCCTCTGGACCCGCC 1345
Db 78284 AGTGTGGTGTGGGACCACTGTGGCCAGAGGGGCAAGAAACAGGCCCTCTGGACCCGCC 78225
Qy 1346 ATCCCTGCGGTGTCCTCCCTCCGCCACCTGGCTGGCCAGCCCGGCTCTCCCTCCTTA 1405
Db 78224 ATCCCTGCGGTGTCCTCCCTCCGCCACCTGGCTGGCCAGCCCGGCTCTCCCTCCTTA 78165
Qy 1406 GCGCCCTCGGGGACCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1465
Db 78164 GCGCCCTCGGGGACCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 78105
Qy 1466 GGCCTGAGGACCACTGAGCTGAGTGGCTCTGCTGGGGGGGAGCACTGGCGGAG 1525
Db 78104 GGCCTGAGGACCACTGAGCTGAGTGGCTCTGCTGGGGGGGAGCACTGGCGGAG 78045
Qy 1526 CTGCGCGCTTGT 1585
Db 78044 CTGCGCGCTTGT 77985
Qy 1586 CACCACACATGAACATGACCACTTACCACTTACCACTTACCACTTACCACTTACCACT 1645
Db 77984 CACCACACATGAACATGACCACTTACCACTTACCACTTACCACTTACCACTTACCACT 77925
Qy 1646 CCAAGGCCCAACCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1705
Db 77924 CCAAGGCCCAACCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 77865
Qy 1706 CTAATTAATTGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1765
Db 77864 CTAATTAATTGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 77805
Qy 1766 ATTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 1825
Db 77804 ATTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 77745
Qy 1826 GTT 1828
Db 77744 GTT 77742

RESULT 14
AC007334/c 170791 bp DNA linear PRI 13-MAR-2003
LOCUS Homo sapiens chromosome 16 clone RP11-147B17, complete sequence.
DEFINITION AC007334
AC007334.8 GI:28933524
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 170791)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
TITLE Alamos National Laboratory.
JOURNAL Direct Submission
AUTHORS Unpublished
2 (bases 1 to 170791)
Bruce, D., Mande, M., Doggett, N., Munk, C., Saunders, B., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, U., Malbale, M., Bussod, W., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
3 (bases 1 to 170791)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 170791)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 13, 2003 this sequence version replaced gi:15284265.
COMMENT Draft Sequence Produced By DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www-sbnc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated total Number of Errors is 0.
Location/Qualifiers
1. 170791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-147B17"
ORIGIN
Query Match 42.1%; Score 783; DB 9; Length 170791;
Best Local Similarity 100.0%; Pred. NO. 1.2e-131;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGCTCCCTCCCTCCGTCGAGCAAGTCAAGTCCGCCCCGCACTCCCAATCCCACTCGA 1105
DB 28069 GGCTCCCTCCCTCCGTCGAGCAAGTCAAGTCCGCCCCGCACTCCCAATCCCACTCGA 28010
QY 1106 TCTGCTCCCAATGAGCCGGAAGCATCAATCCCAACCGAAGCCCGGAGGCTGAC 1165
DB 28009 TCTGCTCCCAATGAGCCGGAAGCATCAATCCCAACCGAAGCCCGGAGGCTGAC 27950
QY 1166 CCGGCTCCCTCCCTCCGTCGAGCAAGTCCGCCCCGCACTCCCAATCCCACTCGA 1225
DB 27949 CCGGCTCCCTCCCTCCGTCGAGCAAGTCCGCCCCGCACTCCCAATCCCACTCGA 27890
QY 1226 CTCGGGGGCAACCCAGAGCGAGCACTTTGTGAGTCCCGCAAGGCGCCAGGGCAAG 1285
DB 27889 CTCGGGGGCAACCCAGAGCGAGCACTTTGTGAGTCCCGCAAGGCGCCAGGGCAAG 27830
QY 1286 AGTGTGGGTGTGGGCACTGTGAGCCAGAGGGGCAAGAAACAAGCCCTCTGGAACCCGCC 1345
DB 27829 AGTGTGGGTGTGGGCACTGTGAGCCAGAGGGGCAAGAAACAAGCCCTCTGGAACCCGCC 27770
QY 1346 ATCCCTGGGTGTGGGCACTGTGAGCCAGAGGGGCAAGAAACAAGCCCTCTGGAACCCGCC 1405
DB 27769 ATCCCTGGGTGTGGGCACTGTGAGCCAGAGGGGCAAGAAACAAGCCCTCTGGAACCCGCC 27710
QY 1406 GCCCTCCCTCGGGGCAACAAGCAACAAGCAACGAGCCCAAGAGAGCCAGAGGGCTGCGG 1465

DB 27709 GCCCTCCCTCGGGGCAACAAGCAACAAGCAACGAGCCCAAGAGAGCCAGAGGGCTGCGG 27650
QY 1466 GGCTGTGAGGCAACCACTGGCTTCAAGTGGCCCTGCTCTGGGGGGGAGCACTGGGGAG 1525
DB 27649 GGCTGTGAGGCAACCACTGGCTTCAAGTGGCCCTGCTCTGGGGGGGAGCACTGGGGAG 27590
QY 1526 CTGCGGCTCTGGGTGTGTGTGAGAGCCAGAGCCGGGCAAGCCGCTCCAGAGATGAGCAC 1585
DB 27589 CTGCGGCTCTGGGTGTGTGTGAGAGCCAGAGCCGGGCAAGCCGCTCCAGAGATGAGCAC 27530
QY 1586 CACCAACCATGAAACATCAACCACTTACCTTACCAAGATGAGAGCCCTCC 1645
DB 27529 CACCAACCATGAAACATCAACCACTTACCTTACCAAGATGAGAGCCCTCC 27470
QY 1646 CCAAGGCCCCCACTGTGCATATGAGAGCCCAACCCCGGACACCAAGGCACTTATAT 1705
DB 27469 CCAAGGCCCCCACTGTGCATATGAGAGCCCAACCCCGGACACCAAGGCACTTATAT 27410
QY 1706 CTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1765
DB 27409 CTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 27350
QY 1766 ATTAGCTACCTTACATGTAAGATCTATGAAACACAGAACTTATTTATTTATTTAT 1825
DB 27349 ATTAGCTACCTTACATGTAAGATCTATGAAACACAGAACTTATTTATTTATTTATTTAT 27290
QY 1826 GTT 1828
DB 27289 GTT 27287
RESULT 15
AC145238 178790 bp DNA linear HTG 15-AUG-2003
LOCUS AC145238
DEFINITION Pan troglodytes clone CH251-568P19, WORKING DRAFT SEQUENCE, 25
ordered pieces.
AC145238
VERSION AC145238.2 GI:33667125
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 178790)
REFERENCE
AUTHORS Antonellis, A., Ayala, K., Beckstrom-Stenberg, S. M., Benjamin, B.,
Blakesley, R. W., Bouffierd, G. G., Brinkley, C., Brooks, S., Cartaga, K.,
Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-D., Hu, P.,
Hurle, B., Idol, J. R., Karlins, E., Kwong, P., Lalic, P., Lee-Lin, S.-Q.,
Legardi, R., Maduro, Q. L., Maduro, V. B., Margulies, E. H., Masello, C.,
Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnov, M. E.,
Prasad, A., Redix-Dugue, N., Schandler, K., Schneider, M. G., Shah, K.,
Sison, C., Stantirip, S., Thomas, J. W., Thomas, P. D., Tsipouri, V.,
Vogt, J. L., Wetherby, K. D., Young, A. and Green, E. D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178790)
Green, E. D.
AUTHORS Direct Submission
TITLE Submitted (18-JUN-2003) NIH Intramural Sequencing Center, 8717
JOURNAL Govemont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 178790)
REFERENCE
AUTHORS Green, E. D.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
Govemont Circle, Gaithersburg, MD 20877, USA
On Aug 15, 2003 this sequence version replaced gi:31880081.
COMMENT ----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@ngri.nih.gov
----- Project Information

Center project name: esk
Center clone name: 568P19

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171494 bases at least Q40
Consensus quality: 174120 bases at least Q30
Consensus quality: 175495 bases at least Q20
Insert size: 149000; agarose-fp
Insert size: 176390; sum-of-contigs
Quality coverage: 10.92x in Q20 bases; agarose-fp
Quality coverage: 9.22x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

1      2604: contig of 2604 bp in length
*
* 2605      2704: gap of unknown length
*
* 2705      7790: contig of 5086 bp in length
*
* 7791      7890: gap of unknown length
*
* 7891      12833: contig of 4943 bp in length
*
* 12834      12933: gap of unknown length
*
* 12934      14664: contig of 1731 bp in length
*
* 14665      14765: gap of unknown length
*
* 14766      32237: contig of 17473 bp in length
*
* 32238      32337: gap of unknown length
*
* 32338      35356: contig of 3019 bp in length
*
* 35357      35456: gap of unknown length
*
* 35457      39292: contig of 3836 bp in length
*
* 39293      39392: gap of unknown length
*
* 39393      41210: contig of 1818 bp in length
*
* 41211      41310: gap of unknown length
*
* 41311      43762: contig of 2452 bp in length
*
* 43763      43862: gap of unknown length
*
* 43863      46766: contig of 2904 bp in length
*
* 46767      46866: gap of unknown length
*
* 46867      53405: contig of 6539 bp in length
*
* 53406      53505: gap of unknown length
*
* 53506      55334: contig of 1829 bp in length
*
* 55335      55434: gap of unknown length
*
* 55435      69095: contig of 13661 bp in length
*
* 69096      69195: gap of unknown length
*
* 69196      78811: contig of 9616 bp in length
*
* 78812      78911: gap of unknown length
*
* 78913      89715: contig of 10804 bp in length
*
* 89716      89815: gap of unknown length
*
* 89816      97259: contig of 7444 bp in length
*
* 97260      97359: gap of unknown length
*
* 97360      99238: contig of 1879 bp in length
*
* 99239      99338: gap of unknown length
*
* 99339      101324: contig of 1986 bp in length
*
* 101325      101424: gap of unknown length
*
* 101425      112066: contig of 10644 bp in length

```

FEATURES

source

```

* 112069      112168: gap of unknown length
*
* 112169      118679: contig of 6511 bp in length
*
* 118680      118779: gap of unknown length
*
* 118780      120680: contig of 1901 bp in length
*
* 120681      120780: gap of unknown length
*
* 120781      135732: contig of 14952 bp in length
*
* 135733      135832: gap of unknown length
*
* 135833      138723: contig of 2891 bp in length
*
* 138724      138823: gap of unknown length
*
* 138824      163652: contig of 24829 bp in length
*
* 163653      163752: gap of unknown length
*
* 163753      178790: contig of 15038 bp in length.
*
* Location/Qualifiers
*   1. 178790
*     /organism="Pan troglodytes"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:9598"
*     /clone="CH251-568P19"
*     /clone_11b="CH251"
*   1. 90615
*     /note="clone overlaps with GenBank Accession Number
*     AC145175 clone CH251-32B1 (center project name esj)"
*   1. 2604
*     /note="assembly_fragment"
*     clone_end:SP6
*     vector_side:left"
*   2705. 7790
*     /note="assembly_fragment"
*     /note="assembly_fragment"
*     7891. 12833
*     /note="assembly_fragment"
*     12934. 14664
*     /note="assembly_fragment"
*     14765. 32237
*     /note="assembly_fragment"
*     32338. 35356
*     /note="assembly_fragment"
*     35457. 39292
*     /note="assembly_fragment"
*     /note="assembly_fragment"
*     39393. 41210
*     /note="assembly_fragment"
*     41311. 43762
*     /note="assembly_fragment"
*     43863. 46766
*     /note="assembly_fragment"
*     46867. 53405
*     /note="assembly_fragment"
*     53506. 55334
*     /note="assembly_fragment"
*     55435. 69095
*     /note="assembly_fragment"
*     69196. 78811
*     /note="assembly_fragment"
*     78912. 89715
*     /note="assembly_fragment"
*     89816. 97259
*     /note="assembly_fragment"
*     97360. 99238
*     /note="assembly_fragment"
*     99339. 101324
*     /note="assembly_fragment"
*     101425. 112066
*     /note="assembly_fragment"
*     112169. 118679
*     /note="assembly_fragment"
*     118780. 120680
*     /note="assembly_fragment"
*     120781. 135732
*     /note="assembly_fragment"
*     /note="assembly_fragment"
*     135833. 138723
*     /note="assembly_fragment"
*     138824. 178790
*     /note="clone overlaps with GenBank Accession Number
*     AC145174 clone CH251-260A1 (center project name esj)"

```

misc_feature 138824..163652 /note="assembly_fragment"
misc_feature 163753..178790 /note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 41.7%; Score 775; DB 2; Length 178790;
Best Local Similarity 99.4%; Pred. No. 3,4e-130;
Matches 778; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1046 GGCTCCCTCCCTCCGCGCCCAAGTCAAACTGCCCCCGCACTCCCAATCCACTCGA 1105
DB 65319 GGCTCCCTCCCTCCGCGCCCAAGTCAAACTGCCCCCGCACTCCCAATCCACTCGA 65378
QY 1106 TCTGCTCCCAAGACCGGAAGCCATCCATCCACACCGAAAGCCCAAGGCGTGGAC 1165
DB 65379 TCTGCTCCCAAGACCGGAAGCCATCCATCCACACCGAAAGCCCAAGGCGTGGAC 65438
QY 1166 CCGGCTCTCTCCCTCCCTCTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGG 1225
DB 65439 CCGGCTCTCTCCCTCCCTCTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGG 65498
QY 1226 CTCGCGGGGCAACCGAGACGGAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAG 1285
DB 65499 CTCGCGGGGCAACCGAGACGGAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAG 65558
QY 1286 AGTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGAGCCGCGC 1345
DB 65559 AGTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGAGCCGCGC 65618
QY 1346 ATCCCTGCGGTGTGTCCTCCCTCCGCACTGCTGCTCCAGCCCGGCTCTCTCCCTCTA 1405
DB 65619 ATCCCTGCGGTGTGTCCTCCCTCCGCACTGCTGCTCCAGCCCGGCTCTCTCCCTCTA 65678
QY 1406 GCCCCTCTCGGGCAAGAAAGCAAGAGCAACGAGCCAGAGAGCCAGAGGCTGCGG 1465
DB 65679 GCCCCTCTCGGGCAAGAAAGCAAGAGCAACGAGCCAGAGAGCCAGAGGCTGCGG 65738
QY 1466 GGCCTGAGGCAACCACTGGCTCAAGTGGCCCTGTCTGTGGGGCGGAGCACTGCGGAG 1525
DB 65739 GGCCTGAGGCAACCACTGGCTCAAGTGGCCCTGTCTGTGGGGCGGAGCACTGCGGAG 65798
QY 1526 CTGCCCCCTTTGGTGTGTATGAGAGCCAGGCCGCGGCAAGGCTCCAGAGCATGAGCAC 1585
DB 65799 CTGCCCCCTTTGGTGTGTATGAGAGCCAGGCCGCGGCAAGGCTCCAGAGCATGAGCAC 65858
QY 1586 CACCACCAACCATGACATCAACCATTTACCACTTTCTACAGACATAGAGCCCTCC 1645
DB 65859 CACCACCAACCATGACATCAACCATTTACCACTTTCTACAGACATAGAGCCCTCC 65918
QY 1646 CCAGGGGCCCAACCTGCAATGAAAGACCCCAACCCCGACACCAAGGCAATTTATT 1705
DB 65919 CCAGGGGCCCAACCTGCAATGAAAGACCCCAACCCCGACACCAAGGCAATTTATT 65978
QY 1706 CTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1765
DB 65979 CTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 66038
QY 1766 ATTAGCTAGCTCAATGATAGATCTATGAAACAGAACTAACTTTATTTATTTAT 1825
DB 66039 ATTAGCTAGCTCAATGATAGATCTATGAAACAGAACTAACTTTATTTATTTAT 66098
QY 1826 GTT 1828
DB 66099 GTT 66101

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM, nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 1305 Seconds
(without alignments)
7477.911 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gaattcgccctctataacg.....aaaaaaaaaagcgccgc 1859

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseq11980s:*
2: geneseq11990s:*
3: geneseq12000s:*
4: geneseq12001as:*
5: geneseq12001bs:*
6: geneseq12002as:*
7: geneseq12002bs:*
8: geneseq12003as:*
9: geneseq12003bs:*
10: geneseq12003cs:*
11: geneseq12003ds:*
12: geneseq12004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1859	100.0	1859	10	AAL51511 Human Nkd
2	1434.8	77.2	1438	3	AAAG3925 DNA encod
3	1221.4	65.7	1731	3	AAAG3924 DNA encod
4	1078.4	58.0	1416	4	AAH28343 Nucleotid
5	1066.6	57.4	1401	10	AAL51512 Mouse Nkd
6	783	42.1	2379	4	AAL03214 Human rep
7	783	42.1	2379	4	AAL03216 Human rep
8	590	31.7	590	10	AAL51525 Human Nkd
9	455.8	24.5	1561	5	AAAG3928 DNA encod
10	418.2	22.5	427	9	ACH21474 Human adu
11	284.2	15.3	1743	5	AAAG78752 DNA encod
12	284.2	15.3	1743	5	AAAG92081 DNA encod
13	284.2	15.3	1743	10	ADG11966 Human nov
14	272.4	14.7	1966	4	AAAG27117 cDNA enco
15	272.4	14.7	1966	10	ADG83295 Human cDN
16	252	13.6	1285	3	AAAG3926 DNA encod
17	245.4	13.2	1885	10	ADG82455 Leukaemia
18	215	11.6	215	10	AAL51526 Human Nkd
19	207	11.1	1959	12	ADG22400 Human sof
20	206.8	11.1	1307	3	AAAG3927 DNA encod
21	148	8.0	148	10	AAL51522 Human Nkd

22	137	7.4	598	12	ACH74795 Human gen
23	137	7.4	1950	6	ABQ73820 Human col
24	134	7.2	134	10	AAL51518 Human Nkd
25	133	7.2	133	12	ACH88495 Human gen
26	128	6.9	128	10	AAL51524 Human Nkd
27	120	6.5	120	10	AAL51514 Human Nkd
28	109	5.9	639	8	ACD05637 cDNA enco
29	107	5.8	107	10	AAL51520 Human Nkd
30	96	5.2	96	10	AAL51521 Human Nkd
31	89	4.8	554	12	ACH74420 Human gen
32	86.4	4.6	179	12	ACH88120 Human gen
33	85	4.6	85	10	AAL51523 Human Nkd
34	72.4	3.9	1817	6	ABK34945 Human cDN
35	70.2	3.8	2723	4	AAK75944 Human imm
36	68.4	3.7	1264	12	ADG24765 Dog TRAIL
37	67	3.6	67	10	AAL51519 Human Nkd
38	65.6	3.5	670	6	ABV73487 Human cyt
39	62.8	3.4	551	10	ADG81081 Leukaemia
40	60.6	3.3	2000	8	ADA71938 Rice gene
41	58.8	3.2	707	6	ABX04893 Corus BP
42	58.2	3.1	1163	12	ADG24767 Cat TRAIL
43	57.4	3.1	510	6	ABV73485 Human cyt
44	57.2	3.1	417	6	ABV73496 Human cyt
45	56.8	3.1	114955	2	AAK53491 Human ade

ALIGNMENTS

RESULT 1	AAL51511	24-APR-2003	(first entry)
ID	AAL51511	standard; DNA; 1859 BP.	
XX	AC	AAL51511;	
XX	DT	24-APR-2003	(first entry)
XX	DE	Human Nkd (hNkd)	protein coding sequence.
XX	KM	Human; gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;	
XX	KW	colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.	
XX	OS	Homo sapiens.	
XX	PH	Key	Location/Qualifiers
FT	FT	CDS	224..1636
FT	FT		/*tag= a
FT	FT		/product= "Human Nkd protein"
XX	PN	W0200292832-A2.	
XX	PD	21-NOV-2002.	
XX	PF	27-NOV-2001; 2001MO-US044092.	
XX	PR	27-NOV-2000; 2000US-0252884P.	
XX	PR	16-MAY-2001; 2001US-0291109P.	
XX	PR	01-OCT-2001; 2001US-0325571P.	
XX	PA	(CHIR) CHIRON CORP.	
XX	PI	Rohan M, Chan V, Yan D;	
XX	DR	WPI; 2003-129303/12.	
XX	DR	P-PSDB; AAO16389.	
XX	PT	New human and non-human primate homologues of Nkd protein, and Nkd genes,	
XX	PT	useful for treating cancer involving aberrant Wnt signaling, e.g. colon	
XX	PS	cancer, head and neck cancer, ovarian cancer, or breast cancer.	
XX	PS	Claim 4; Fig 3; 99pp; English.	
XX	CC	The invention comprises the amino acid and coding sequence of the human	

CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
CC protein of the invention
XX

SQ Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 Other;

Query Match 100.0%; Score 1859; DB 10; Length 1859;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATTCCTCCCTTAATATGAGCTCACTATAGGAGCAAGAGTGTATCAACAGCAGATACG 60
Db 1 GAATTCCTCCCTTAATATGAGCTCACTATAGGAGCAAGAGTGTATCAACAGCAGATACG 60
QY 61 CGGGAGTCGGGCGCGCGCGAGCGGCGAGAGGCGCTCCCGGCGCGCTTCGGGCTCCG 120
Db 61 CGGGAGTCGGGCGCGCGCGAGCGGCGAGAGGCGCTCCCGGCGCGCTTCGGGCTCCG 120
QY 121 CTGGGCTCGGGGGGCTGCTTGGGAGAGAGAGAGCCAAAGGAGGCGCCAGAGCGCGGGCC 180
Db 121 CTGGGCTCGGGGGGCTGCTTGGGAGAGAGAGAGCCAAAGGAGGCGCGCCAGAGCGCGGGCC 180
QY 181 GGGGCGATGGCTTAGGAGCGCTCCGGCGCGCGAGCCCGCAGCATGGGAGAACTTCACTC 240
Db 181 GGGGCGATGGCTTAGGAGCGCTCCGGCGCGCGAGCCCGCAGCATGGGAGAACTTCACTC 240
QY 241 CAAGCCGCGCGCGCTGTGCAAGCGCAGAGAGAGCCCGAGAGTGAACGCTTTCGCGTAG 300
Db 241 CAAGCCGCGCGCGCTGTGCAAGCGCAGAGAGAGCCCGAGAGTGAACGCTTTCGCGTAG 300
QY 301 CGCTGCTGGGCTCGGAGAGGCGCATCGAGAGTGAATCGGAGAGAGCGCTGCGCGGCGG 360
Db 301 CGCTGCTGGGCTCGGAGAGGCGCATCGAGAGTGAATCGGAGAGAGCGCTGCGCGGCGG 360
QY 361 TGTCTCGGAGACCCGAGAGCTGCGGTTGGCGGACACCATAGGCGGAGACCCCGGAGCT 420
Db 361 TGTCTCGGAGACCCGAGAGCTGCGGTTGGCGGACACCATAGGCGGAGACCCCGGAGCT 420
QY 421 CGTGGGCGACGCTGTTGAGAGACACGCTCAAGGAGAGAGAGAGAGCGATTTTCGCTTGA 480
Db 421 CGTGGGCGACGCTGTTGAGAGACACGCTCAAGGAGAGAGAGAGAGCGATTTTCGCTTGA 480
QY 481 AGTGGCCCTGCTCTCTGAAGAGACTGAACGGCTGGGAGCGGAGATGAAGAGATGGA 540
Db 481 AGTGGCCCTGCTCTCTGAAGAGACTGAACGGCTGGGAGCGGAGATGAAGAGATGGA 540
QY 541 GAGAGTAGAGCAACCTTGCCAGAGCTCCAAAGACAGCTGAAGTTTGAAGAGCTCCAGTG 600
Db 541 GAGAGTAGAGCAACCTTGCCAGAGCTCCAAAGACAGCTGAAGTTTGAAGAGCTCCAGTG 600
QY 601 CGACGTGTCAATGAGAGAGAGACAGCCGAGAGAGTGAACCTTCAACCTGTATGAATTGA 660
Db 601 CGACGTGTCAATGAGAGAGAGACAGCCGAGAGAGTGAACCTTCAACCTGTATGAATTGA 660
QY 661 CAACAAAGGAGAGTCAACCGAGAGAGCATCAACAGCTGTGACACCATCTATAGAGT 720
Db 661 CAACAAAGGAGAGTCAACCGAGAGAGCATCAACAGCTGTGACACCATCTATAGAGT 720
QY 721 GGTGAGCTCTCTCTCAACCACTCCCAACATCCAGCAAGATGTGCGGGTAAAGCTCAC 780
Db 721 GGTGAGCTCTCTCTCAACCACTCCCAACATCCAGCAAGATGTGCGGGTAAAGCTCAC 780
QY 781 CGTGGCCCGCATGAGAGAGAGCAAGAGAGAGCGTCTTGTCAATCAAGCTGAACCTGCA 840
Db 781 CGTGGCCCGCATGAGAGAGAGCAAGAGAGAGCGTCTTGTCAATCAAGCTGAACCTGCA 840
QY 841 GAGGCGAAGGCGCCGAGAGAGAGCAACCAAGCCCATGAGAGACTTGGGAGCTGGAGAGAA 900
Db 841 GAGGCGAAGGCGCCGAGAGAGAGCAACCAAGCCCATGAGAGACTTGGGAGCTGGAGAGAA 900
QY 901 GCACGAGCCCGCTCAGGTTCCAGGGTGAACAGCGGCTGAGAGAGTCTGCTGTACCA 960
```

```
Db 901 GCACGAGCCCGCTCAGGTTCCAGGGTGAACAGCGGCTGAGAGAGTCTGCTGTACCA 960
QY 961 CCAATGCTGTATGATGAGAACTCGAGAGAGAGAAACCACTACTTATGATTCGCGGGATAGA 1020
Db 961 CCAATGCTGTATGATGAGAACTCGAGAGAGAGAGAAACCACTACTTATGATTCGCGGGATAGA 1020
QY 1021 AAACCTACAGTCCCAATTGAGGCTCGGCTCCCTTCGCTGAGCCAGAAAGTCAGAACTGCC 1080
Db 1021 AAACCTACAGTCCCAATTGAGGCTCGGCTCCCTTCGCTGAGCCAGAAAGTCAGAACTGCC 1080
QY 1081 CCCCAGCACTCCCAATCCCATGATCTGCTCCCATGAGCCGAGAGCCATCCATGCC 1140
Db 1081 CCCCAGCACTCCCAATCCCATGATCTGCTCCCATGAGCCGAGAGCCATCCATGCC 1140
QY 1141 ACACCGAAGAGCCCAAGGCGTGAACCGGCTCTCTTCACTTCTCTTGAACCCCAATGCC 1200
Db 1141 ACACCGAAGAGCCCAAGGCGTGAACCGGCTCTCTTCACTTCTCTTGAACCCCAATGCC 1200
QY 1201 CAAGGTCTCAGAGCTCCAGCAAGGCTCCGAGGACCCAGAGACGAGAGCAACTTGT 1260
Db 1201 CAAGGTCTCAGAGCTCCAGCAAGGCTCCGAGGACCCAGAGACGAGAGCAACTTGT 1260
QY 1261 GAGGTCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAAGTGTGGGCAAGTGTGGGCAAG 1320
Db 1261 GAGGTCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAAGTGTGGGCAAGTGTGGGCAAG 1320
QY 1321 AAACAGAGCCCTCTGAGAGACCCGCTCATCTCGGAGGTATCCCTCCGCGCCACCTGGCTGC 1380
Db 1321 AAACAGAGCCCTCTGAGAGACCCGCTCATCTCGGAGGTATCCCTCCGCGCCACCTGGCTGC 1380
QY 1381 CAGCCGAGCCCTCTCCCTCTAGAGCCCTCTGAGGCAAGAGAGCAACGAGCCGAGC 1440
Db 1381 CAGCCGAGCCCTCTCTCCCTCTAGAGCCCTCTGAGGCAAGAGAGCAACGAGCCGAGC 1440
QY 1441 CAAAGAGAGCAAGAGGCTGCCGAGGCTGCAAGAGCAACTGAGCTCAGGTGAGCTGT 1500
Db 1441 CAAAGAGAGCAAGAGGCTGCCGAGGCTGCAAGAGCAACTGAGCTCAGGTGAGCTGT 1500
QY 1501 CTTGGGCGGAGAGCACTGCGGAGAGCTGCGGCTTGGTGTGTATGAGAGCCAGGCGG 1560
Db 1501 CTTGGGCGGAGAGCACTGCGGAGAGCTGCGGCTTGGTGTGTATGAGAGCCAGGCGG 1560
QY 1561 GCAAGCGGTCAAGAGCATGAGAGCAACCAACCATGAACATCAACCATTTACCA 1620
Db 1561 GCAAGCGGTCAAGAGCATGAGAGCAACCAACCATGAACATCAACCATTTACCA 1620
QY 1621 CTTCTACAGACATGAGAGCCCTCCAGAGGCGCCACCTGCAATGAGAGAGCCCAAC 1680
Db 1621 CTTCTACAGACATGAGAGCCCTCCAGAGGCGCCACCTGCAATGAGAGAGCCCAAC 1680
QY 1681 CCGGACACCAAGGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1740
Db 1681 CCGGACACCAAGGCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1740
QY 1741 TAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1800
Db 1741 TAAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1800
QY 1801 CACGAGACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1859
Db 1801 CACGAGACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1859
```

RESULT 2
AAA63925
ID AAA63925 standard; DNA; 1438 BP.

XX AAA63925;

XX 04-DEC-2000 (first entry)

XX DNA encoding protein related to Drosophila naked cuticle polypeptide.

XX Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
KW gene therapy; Nkd defect; cancer; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 8..1420
FT /tag= a
FT /product= "naked cuticle polypeptide"
XX
XX MO200049034-A1.
XX
XX 24-AUG-2000.
XX
XX 17-FEB-2000; 2000MO-US004188.
XX
XX 17-FEB-1999; 99US-0120646P.
XX
XX (STRD) UNIV LELAND STANFORD JUNIOR.
XX
XX Scott M, Zeng W, Wharton K;
XX
XX MPI: 2000-571967/53.
XX P-PSDB; AAB08216.
XX
XX An isolated nucleic acid molecule useful for analyzing (genetic
PT predilection to) a disease state and for therapeutic purposes e.g.
PT treatment of cancer comprises a sequence encoding a naked cuticle
PT protein.
XX
XX Claim 3; Page 46-48; 58pp; English.
XX
XX
XX The present sequence encodes a protein related to the Drosophila Nkd
CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide
CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the
CC regulation of Wnt signal potency, duration or distribution. The Nkd
CC polynucleotide can be used for identifying homologous or related
CC proteins, to modulate the expression or function of Nkd polypeptides, and
CC in studying associated physiological pathways. Nkd polynucleotides can
CC also be used in gene therapy to treat disorders associated with Nkd
CC defects. They may also be used for therapeutic purposes e.g. treatment of
CC cancer
XX
SQ Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;
Query Match 77.2%; Score 1434.8; DB 3; Length 1438;
Best Local Similarity 99.9%; Pred. No. 3.9e-277;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 217 CCCAGCATGGGAACTTCACTCAAGCCGCGCCGTGCAAGCGCAGGAGAGCC 276
DB 1 CCCAGCATGGGAACTTCACTCAAGCCGCGCCGTGCAAGCGCAGGAGAGCC 60
OY 277 GGAAGGTGACGCTTGCCTGAGCCGCTGCTGAGGCTCGGAAGGACATGAGAGT 336
DB 61 GGAAGGTGACGCTTGCCTGAGCCGCTGAGGCTCGGAAGGACATGAGAGT 120
OY 337 CGGGAACAGCCGCTGCGCCGCGCGCTGCTGAGGACCCCGACAGCTGGGTGGCGG 396
DB 121 CGGGAACAGCCGCTGCGCCGCGCGCTGCTGAGGACCCCGACAGCTGGGTGGCGG 180
OY 397 CATAGCCGGAAGACCCCGGAGCTCGTGGGCGAGCTTTGAGAGACAGCTCAGCGAGA 456
DB 181 CATAGCCGGAAGACCCCGGAGCTCGTGGGCGAGCTTTGAGAGACAGCTCAGCGAGA 240
OY 457 AGAGAGAGACGACTTTCGCTGGAAGTGGCCCTGCTTCTGAGAACTGAACGAGCTGG 516
DB 241 AGAGAGAGACGACTTTCGCTGGAAGTGGCCCTGCTTCTGAGAACTGAACGAGCTGG 300
OY 517 CACGGAGATGAGAAAGATGGAGAGTGAAGGAACTCCGCGCAGGCTCCAAAGAGCA 576
|||||

DB 301 CACGGAGATGAGAAAGATGGAGAGTGAAGCAACCTGCCACGCTCCAAAGAGCA 360
OY 577 GCTGAAGTTTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG 636
DB 361 GCTGAAGTTTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGAGACAGCCGGCAGAGTG 420
OY 637 GACCTTCAACCCCTTAAGACTTTTGAACAACAACGCGCAAGGTCAACCCGAGAGACATCACAG 696
DB 421 GACCTTCAACCCCTTAAGACTTTTGAACAACAACGCGCAAGGTCAACCCGAGAGACATCACAG 480
OY 697 CTGCTGACACCATCATATGAGGTGGTGAAGTCTCTCTGCAACCACTCCCAATCAG 756
DB 481 CTGCTGACACCATCATATGAGGTGGTGAAGTCTCTCTGCAACCACTCCCAATCAG 540
OY 757 CAAGATGCTGGGATTAAGTCAACCTGACCCCGATGGAGCCAGAGCAAGAGAGAGCT 816
DB 541 CAAGATGCTGGGATTAAGTCAACCTGACCCCGATGGAGCCAGAGCAAGAGAGAGCT 600
OY 817 CTTGTCAATCAGAGCTGACCTGCAAGAGCCGCAAGCCGCGAGAGCAAGCCCACTGA 876
DB 601 CTTGTCAATCAGAGCTGACCTGCAAGAGCCGCGAGAGCAAGCCCACTGA 660
OY 877 GGAAGCTGGAGCTGGGAG 936
DB 661 GGAAGCTGGAGAGCTGGGAG 720
OY 937 CTTGAGAGAGTGGCTGCTACCAACCATTTGATGAGAGATGAGAGATGAGAGAGAGAGAG 996
DB 721 CTTGAGAGAGTGGCTGCTACCAACCATTTGATGAGAGATGAGAGATGAGAGAGAGAGAG 780
OY 997 CTACTTAAGATCTGGCGGAGTAAAGAACTAAGTCCCAATTTGGGCTGCTCCCTTC 1056
DB 781 CTACTTAAGATCTGGCGGAGTAAAGAACTAAGTCCCAATTTGGGCTGCTCCCTTC 840
OY 1057 CGTGGCCAGAGATGAGAGAGTGGCCCGCCGACCTCCATCCCATGATCTGCTCCCA 1116
DB 841 CGTGGCCAGAGATGAGAGAGTGGCCCGCCGACCTCCATCCCATGATCTGCTCCCA 900
OY 1117 TGAAGCGGAAGCATTCACATCCCAACCGAAAGCCCAAGGCGTGAACCCGAGCTCTCT 1176
DB 901 TGAAGCGGAAGCATTCACATCCCAACCGAAAGCCCAAGGCGTGAACCCGAGCTCTCT 960
OY 1177 CCACTTCTTGAACCCCAATTCGCAAGTCTCAGAGCTCCACAGACGCTCCGAGGAC 1236
DB 961 CCACTTCTTGAACCCCAATTCGCAAGTCTCAGAGCTCCACAGACGCTCCGAGGAC 1020
OY 1237 CCAAGACGGAAGCAAGCATTTGTGAGGTCCCAAGAGCCCAAGAGAGAGTGTGAGTGT 1296
DB 1021 TCAAGACGGAAGCAAGCATTTGTGAGGTCCCAAGAGCCCAAGAGAGAGTGTGAGTGT 1080
OY 1297 GAGGCAAGTGGCCAGAGAGGCAAGAAACAAGCCCTCTGAGAACCCGCTGCGGT 1356
DB 1081 GAGGCAAGTGGCCAGAGAGGCAAGAAACAAGCCCTCTGAGAACCCGCTGCGGT 1140
OY 1357 GTCCCTCTGCGCCCACTGCTGCTGAGCCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 1416
DB 1141 GTCCCTCTGCGCCCACTGCTGCTGAGCCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
OY 1417 GCAAGAAAGCAAGAGACCGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
DB 1201 GCAAGAAAGCAAGAGACCGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
OY 1477 ACCACTGGCTCAAGTGGCCCTGTCTGTGGGGCGGAGACCTGCGGAGCTGCGGCTT 1536
DB 1261 ACCACTGGCTCAAGTGGCCCTGTCTGTGGGGCGGAGACCTGCGGAGCTGCGGCTT 1320
OY 1537 GGTGTGTATGAGAGCAAGGCGCGGAGAGCGGCTGCAAGAGCAAGAGAGCAACCAACCA 1596
DB 1321 GGTGTGTATGAGAGCAAGGCGCGGAGAGCGGCTGCAAGAGCAAGAGAGCAACCAACCA 1380
OY 1597 TGAACATCAACCAACATTAACCAACATTTCTACAGACATAGAGCCCTCCCAAGGAGCC 1654
DB 1381 TGAACATCAACCAACATTAACCAACATTTCTACAGACATAGAGCCCTCCCAAGGAGCC 1438
|||||

```
RESULT 3
ID   AAA63924 standard; DNA, 1731 BP.
AC   AAA63924;
DT   04-DEC-2000 (first entry)
XX
DE   DNA encoding protein related to Drosophila naked cuticle polypeptide.
XX
KM   Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
XX   gene therapy; Nkd defect; cancer; se.
XX
OS   Mus musculus.
XX
FH   Key Location/Qualifiers
FT   CDS 140..1555
FT           /*tag= a
FT           /product= "naked cuticle polypeptide"
FT
XX
XX   MO200049034-A1.
XX
XX   24-AUG-2000.
XX
XX   17-FEB-2000; 2000WO-US004188.
XX
XX   17-FEB-1999; 99US-0120646P.
XX
XX   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
XX   Scott M, Zeng W, Wharton K;
XX
XX   WPI; 2000-571967/53.
XX   DR P-PSDB; AAB08215.
XX
XX
XX   An isolated nucleic acid molecule useful for analyzing (genetic
XX   predilection to) a disease state and for therapeutic purposes e.g.
XX   treatment of cancer comprises a sequence encoding a naked cuticle
XX   protein.
XX
XX   Claim 3; Page 43-45; 58pp; English.
XX
XX
XX   The present sequence encodes a protein related to the Drosophila Nkd
XX   (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
XX   gene whose expression is induced by Wnt signalling. The Nkd polypeptide
XX   acts to antagonize Wnt signalling. Nkd may link ion fluxes to the
XX   regulation of Wnt signal potency, duration or distribution. The Nkd
XX   polynucleotides can be used for identifying homologous or related
XX   proteins, to modulate the expression or function of Nkd polypeptides, and
XX   in studying associated physiological pathways. Nkd polynucleotides can
XX   also be used in gene therapy to treat disorders associated with Nkd
XX   defects. They may also be used for therapeutic purposes e.g. treatment of
XX   cancer.
XX
XX
XX   Sequence 1731 BP; 414 A; 521 C; 504 G; 292 T; 0 U; 0 Other;
XX
XX
XX   Query Match 65.7%; Score 1221.4; DB 3; Length 1731;
XX   Best Local Similarity 83.9%; Pred. No. 2e-234;
XX   Matches 1460; Conservative 0; Mismatches 256; Indels 25; Gaps 6;
XX
XX
XX   82 CGCGCGCAGAGCCGCTCCGCGCGCGCTCGGCTCGGCTCGGCGGCTGCTTCG 141
XX   |||||
XX   6 CCGGCGCGCGAGCCGCTCCGCGCGCGCTAAAGGCTGCGCTCGGCGCGGAGATG--TGA 63
XX   |||||
XX   142 GAGAGAGAGAGCAGAGGAGAGCGCCAGGCGCGCGCGCGCGCGCATAGCTTAAGAGAGC 201
XX   |||||
XX   64 GAGAGAGAGCGAGCA--GAGTGGCGCGGAGTGGCGCGCGCGAGCAGCATAGCTTAAGAGAGCT 121
XX   |||||
XX   202 TCCTCGCGCGCGCAGCCCGCAGCATAGGAGAACTTCACTCCAGCGCGCGCGCTGTGCA 261
XX   |||||
XX   122 CCGCGCGCGC---GCCCGCAGCATGGGGAACCTTCACTCGAAGCGCGCGCGCTGTGCA 177
XX   |||||
```

```
QY 262 GCGCAGGAGAGAGCCCGGAGAGTGA CAGCTTCCGCGTGAAGCGCTGCGGCTCGGAAAGG 321
DB |||||
QY 178 GCGCAGGAGAGAGCCCGGAGAGTGA CAGCTTGTGTAAGCGCTGCTTGGCAGAGAAAG 237
DB |||||
QY 322 CATCGAGAGTGAATCGGAGAGACAGCGCTGCCGGGCGGTGTCTCGGAGACCCGACAGCT 381
DB |||||
QY 238 CATCGAGAGTGAATCGGAGAGAGCGCTGTCCAGGAGAGGGTCTCAAGACCCCGTCAAGCT 297
DB |||||
QY 382 GCGGTTGGCGGAGCAGATAGCGGAGACACCCGGAGCTGTGGGCGACGTGTGAAGA 441
DB |||||
QY 298 GAGATTGGCAGAGCACTTGTGTGTCAGGACACTCGGGAACCTGTGGGTCACTTCTAAGA 357
DB |||||
QY 442 CAGGCTCAGGAGAGAGAGAGAGAGAGCACTTCGGCTGGAAGTGGCCCTGCTCTGTGA 501
DB |||||
QY 358 GGCCTCTGGTGAAGAGAGAGAGAGAGAGCACTTCCCTGAAGATGGCCCTGCGCTGAAGA 417
DB |||||
QY 502 GACTGACGGGCTGGGCGACCGAGATGAGAAAGATGAGAGAGTGAAGCGAAACCTGCCC 561
DB |||||
QY 418 GATCGACAGGCTAGGTAGTGAAGATGAGAAAGATGAGAGAGTGAAGCGAAACCTGGCCA 477
DB |||||
QY 562 AGGCTCCAGAGAGAGAGCTGAAGTTTGAAGAGCTTCAGTCCGAGGTCTCATGAGAGAG 621
DB |||||
QY 478 GGCCTCCAGAGAGAGAGCTCAAGTTTGAAGAGCTTACAGTGTGATGTCTCTGTGAGAGAG 537
DB |||||
QY 622 CAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 681
DB |||||
QY 538 CAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597
DB |||||
QY 682 AGAGAGATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 741
DB |||||
QY 598 TGAAGAGATTAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
DB |||||
QY 742 CTCCCAAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801
DB |||||
QY 658 TTCCCAAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
DB |||||
QY 802 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 861
DB |||||
QY 718 GAGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777
DB |||||
QY 862 GACCAAGCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
DB |||||
QY 778 CACCAAGCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
DB |||||
QY 922 CAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 981
DB |||||
QY 838 CAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
DB |||||
QY 982 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1041
DB |||||
QY 898 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957
DB |||||
QY 1042 GCGTGGCTCCCTTCGCTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101
DB |||||
QY 958 ACCGAGATCCCTTCCTTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017
DB |||||
QY 1102 TCGATCTGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1161
DB |||||
QY 1018 TCGGCTTCGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
DB |||||
QY 1162 GAGAGCGGCTCTCTTCCATCTTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1221
DB |||||
QY 1078 GAGAGCGGCTCTCTTCCATCTTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137
DB |||||
QY 1222 ACCGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1281
DB |||||
QY 1138 ACCGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
DB |||||
QY 1282 CAGAGAGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1341
DB |||||
QY 1198 CAGAGAGATGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
DB |||||
```

QY 1342 GCGCATCCCTGGGCTGTCCTCCGCGCAGCTGGCTGCGAGCCGGCCCTCTCCCTC 1401
 DB 1258 GACDACCCTACTGCTCTCTCCCTCTGCGCAGCTGGCGACAGCCCGCTTCTCCAC 1317
 QY 1402 CCTAGCCCCCTGGGCGCAGAAGACAGACCGAGCCAGAGAGAGCCAGAGGCTG 1461
 DB 1318 CTGGGACCCCTGGGCGCAGAAGACAGACCGAGCCAGAGAGAGCCAGAGGCTG 1377
 QY 1462 CCGGGGCTTGCAGGACCACTGGCTTGAAGTGG--CCCTGCTCTGGGGCGGAGCACT 1518
 DB 1378 CCGGGGCTTGCAGGACCCCTGGCTGCGAGGCTCCACCGTCATGGGGCGGAGCAAGT 1437
 QY 1519 GCGGAGCTGCGCCGCTTGGTGTGTATGAGAGCGAGCCGGGCGAGCCGCTCCAGACA 1578
 DB 1438 GAGGAGCTGCTGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1497
 QY 1579 TGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1638
 DB 1498 GGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1557
 QY 1639 CCGCTCCCGAGGCGCCAGCTGCGATATGAGAGAGCCCGGAGCAGCAGCAGCAGT 1698
 DB 1558 CC-----CAGAGGCTGCGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1603
 QY 1699 TATTATTTCTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1758
 DB 1604 TATTATTTCTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1663
 QY 1759 CACTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1818
 DB 1664 CACTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1723
 QY 1819 T 1819
 DB 1724 T 1724
 RESULT 4
 ID AAH28343 standard; DNA; 1416 BP.
 XX
 AC AAH28343;
 XX
 DT 05-SEP-2001 (first entry)
 XX
 DE Nucleotide sequence of a murine mnkd protein.
 XX
 KM Dishevelled-associated protein 1A; DAPIA; mnkd; dishevelled protein;
 XX
 KM Wnt signalling; colon cancer; cancer; 88.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1416
 FT /tag= a
 FT /product= "mnkd"
 XX
 PN W0200144279-A2.
 XX
 PD 21-JUN-2001.
 XX
 PF 05-DEC-2000; 2000MO-US032986.
 XX
 PR 17-DEC-1999; 99US-0172434P.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Yan D, Williams LT;
 XX
 DR WPI; 2001-425440/45.
 DR P-PSDB; AAB84645.
 XX
 PT Novel mammalian dishevelled-associated proteins, mnkd and DAPIA, useful

PT for inhibiting Wnt signaling in mammalian cells, and thus for treating
 PT colon cancer.
 XX
 PS Claim 2; Fig 1; 67pp; English.
 XX
 CC The present sequence encodes a murine mnkd protein. The specification
 CC describes mnkd and dishevelled-associated protein 1A (DAPIA) proteins,
 CC that interact with mammalian dishevelled protein. mnkd is useful for
 CC inhibiting Wnt signalling in mammalian cells, and thus for treating colon
 CC cancer. mnkd is also useful for activating the JNK pathway. mnkd and
 CC DAPIA are also useful for screening drugs that are useful for treating
 CC cancer.
 XX
 SQ Sequence 1416 BP; 339 A; 434 C; 419 G; 224 T; 0 U; 0 Other;
 Query Match 58.0%; Score 1078.4; DB 4; Length 1416;
 Best Local Similarity 85.6%; Pred. No. 7.3e-206;
 Matches 1212; Conservative 0; Mismatches 201; Indels 3; Gaps 1;
 QY 224 ATGGGGAACCTTCACTCCAGAGCCGCGCCGTGTGCAAGCGGAGAGAGAGAGAGAGAGT 283
 DB 1 ATGGGGAACCTTCACTCCAGAGCCGCGCCGTGTGCAAGCGGAGAGAGAGAGAGAGT 60
 QY 284 GACAGCTTGGCGGTGAGCGCTGCTGGCTGGAGGCGATCGAGAGAGAGAGAGAGAGAGAG 343
 DB 61 GACAGCTTGGCGGTGAGCGCTGCTGGAGGCGATCGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 344 CAGCGCTGCGCGGCGGTGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
 DB 121 CAGCGCTGCTCGAGGAG 180
 QY 404 GGAAGCAGCCGAG 463
 DB 181 GGAAGCAGCTCGGAG 240
 QY 464 GAGAGCTTTGGGCGGAG 523
 DB 241 GAGAGCTTTGGGCGGAG 300
 QY 524 GATGAG 583
 DB 301 GATGAG 360
 QY 584 TTGGAAGAGCTCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 643
 DB 361 TTGGAAGAGCTCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 QY 644 ACCCTGATGAGCTTTGAG 703
 DB 421 ACTCTATATGAGCTTTGAG 480
 QY 704 CACAGCATATGAG 763
 DB 481 CATATCATATGAG 540
 QY 764 CTGCGGTAAGAGCTCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823
 DB 541 CTGCGGTAAGAGCTCAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 824 AATCAGAGCTGAGCTGAG 883
 DB 601 AATCAGATCAGATGAG 660
 QY 884 CGAGCTGAG 943
 DB 661 CGAGCTGAG 720
 QY 944 CAGTCTGCTGCTAAG 1003
 DB 721 CAGCAGAGAGCTAAG 780
 QY 1004 GATCTGCGGAG 1063

```

Db      781 GACCTGGCGGGGATAGAGAACTACAGCTCTCAGTTTGACCGGGATCCCTTGCGTGGCC 840
QY      1064 CAGAAATGCAAAATCTGCCCCCGCACTCCAAATCCCACTGATTCGCTTCCATGAGCCG 1123
Db      841 CAAAGTCAGAGCTGCCCCCTCGAATCTCAACCCCACTGCTCTCCCTCCCAAGACCA 900
QY      1124 GAAGCCATCCACATCCCAAGCCCAAGCCCAAGGCGCTCTTCCATTTC 1183
Db      901 GAAGTCGCCACATCCCAACCGGAGGCCCAAGGTGAGACCAAGCTCTTCCACTTC 960
QY      1184 CTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAAGCTCCGAGGCAACCCAGAC 1243
Db      961 CTTGACACCCCAATTTGCAAGGCTCAGAGCTCCAGCAAGGCTCCGAGGCACTCAGAT 1020
QY      1244 GGGAGCAGACATTTTGAGAGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGTGGGCAC 1303
Db      1021 GGGAGCAGACATTTTGAGAGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGTGGGCAC 1080
QY      1304 GTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGCAATCCCTGGCGTGTCCCC 1363
Db      1081 GGGGCCAGAGGTCAAGAAAGCAAGCTCTCACTGTGTACCAACCCCACTACTGTCTCCCC 1140
QY      1364 TCCGCCACCTGGCTGTCAGCCCGCCCTCTCTCCCTCCCTAGCCCCCTTCGGGCAAG 1423
Db      1141 TCTGCCATCTGGCCACCAAGCCCAAGCCCTTCTCCCACTCCGCACTCCGAGGCAAG 1200
QY      1424 AAGCACAAGCAACGAGCCAAAGAGAGCAAGGAGGCTGGGGGCTTCGAGGCACTG 1483
Db      1201 AAACACAAGATCGAGCCAAAGAGCAAGCCAGGCGGAGGCTTCGAGGAGCCCTG 1260
QY      1484 GCTCAGAGTG---CCCTGTCTGGGGGAGAGCACTGCGAGAGTGCCTGCTTGGT 1540
Db      1261 GCTGCAAGAGGCTTCAACCTGTATGGGCGGAGAGAGTGAAGAGTGTGCTGCTGGTG 1320
QY      1541 GTGTATGAGAGCCAGGCGCGGAGCCGCTCCAGACATGAGCAACCAACACCATGAA 1600
Db      1321 GTGTAGCAGAGCCAGGCTGGGAGAGCGCTCCAGAGACACGAACACATCAACGAA 1380
QY      1601 CATCACCACTTATACCACTTATACCAACATAG 1636
Db      1381 CATCACCACTTATATCACTTATATAGCCCTAG 1416

```

RESULT 5

```

ID      AAL51512 standard; DNA; 1401 BP.
XX
AC      AAL51512;
XX
DT      24-Apr-2003 (first entry)
XX
DE      Mouse Nkd protein gene sequence.
XX
KW      Mouse; gene; ds; gene therapy; Nkd; hnkd; Wnt signalling pathway; cancer;
KW      colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
KW      murine.
XX
OS      Mus sp.
XX
PN      WO200292832-A2.
XX
PD      21-NOV-2002.
XX
PF      27-NOV-2001; 2001WO-US044092.
XX
PR      27-NOV-2000; 2000US-0252884P.
PR      16-MAY-2001; 2001US-0291109P.
PR      01-OCT-2001; 2001US-0325571P.
XX
PA      (CHIR ) CHIRON CORP.
XX
PI      Rohan M, Chan V, Yan D;
XX

```

```

DR      WPI; 2003-129303/12.
XX
PT      New human and non-human primate homologues of Nkd protein, and Nkd genes,
PT      useful for treating cancer involving aberrant Wnt signaling, e.g. colon
PT      cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
PS      Example 2; Fig 4; 99p; English.
XX
CC      The invention comprises the amino acid and coding sequence of the human
CC      Nkd protein (hnkd) - a regulator of Wnt signalling pathways. The hnkd DNA
CC      and protein sequences are useful for the treatment of cancer involving
CC      aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC      cancer and breast cancer). The present DNA sequence represents the mouse
CC      Nkd gene
XX
SQ      Sequence 1401 BP; 336 A; 430 C; 416 G; 219 T; 0 U; 0 Other;
Query Match      57.4%; Score 1066.6; DB 10; Length 1401;
Best Local Similarity 85.6%; Pred. No. 1.7e-203;
Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;
QY      224 ATGGGAAATTTCACTTCAAGCCCGCCGCTGTGCAAGGCGAGGAGAGCCCGGAAGT 283
Db      1 ATGGGAAATTTCACTTCAAGCCCGCCGCTGTGCAAGGCGAGGAGAGCCCGGAAGT 60
QY      284 GACAGCTTGGCGGTGAGCGCTGCTGGAGCCCGAAGGAGCATGAGAGTGGAGAG 343
Db      61 GACAGCTTGTCTGTAAAGCTGCTTGGAGCAAGAAAGCATCAAGAGTGGAGAG 120
QY      344 CAGCGCTGCGCGGCGGTGTCTCGGAGACCCCGACAGCTGCGGTGGCGGACCATAGGC 403
Db      121 CAGCGCTGTCAAGCAGCGCTCTCAGAGACCCCGTCAAGTGTGAGTTGGAGGACACTGTGGT 180
QY      404 CGAAGCACCCGGAGAGCTGTGGGCGAGCTGTGAGAGACAGCTCAAGGAGAGAGAG 463
Db      181 CGAAGCAGCTGGGAACTGTGGGTGACATTTAGAGAGGCTCTCGGTGAGAGAGAG 240
QY      464 GACGACTTTGGCTGGAAGTGGCCCTGCTCTGAGAGAGCTGACGGGCTGGGAGCGGA 523
Db      241 GACGACTTCCCTTGAAGAGTGGCTTCCCTGAGAGAGTGCAGAGCTTGAAGTGGGA 300
QY      524 GATGAGAAAGATGAGAGAGTGAAGCAACCTGCCAGGCTTCAAGAAAGACGTTAAG 583
Db      301 GATGAGAAAGATGAGAGAGTGAAGCACTGGCCAGGCTTCAAGAAAGACGTTAAG 360
QY      584 TTGGAAGAGCTCAAGTCCGACGCTGCTCATGAGAGAGAGACCGGCGAGAGTGAACCTTC 643
Db      361 TTGGAAGAGCTTACAGTGTGATGTCTGTGAGAGAGAGACCGGCAAGAGTGAATTTTC 420
QY      644 ACCCTGATGACTTTGACAAACAGGCAAGGTCAACCGAAGAGCATCACAGCTTGGCTG 703
Db      421 ACTCTATATGACTTTCAGCAACATGCAAGTACCCGTGAGAGCATTTACCAAGCTTGGCTG 480
QY      704 CACACATCTATGAGTGTGAGCTCTCTGTCAACCATCTCCCAATCCAGCAAGATG 763
Db      481 CATACATCTATGAGTGTGAGCTCTCTGTGAAACATTTCCCAATCCAGCAAGATG 540
QY      764 CTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGGAGCAAGAGAGAGAGAGAGCTTGTGC 823
Db      541 CTGGGGGTGAAAGCTCACCGTGGCTCTGAGCGGAGCCAGAGTGAAGAGAGAGAGCTTGTTC 600
QY      824 AATCAGGCTGACCTGCAAGCGGAGGCCCGGAGCAGAGACCAAGCCCATGAGAGACTG 883
Db      601 AATCATATCCGATCTGCAAGAGCAAGGCCCGGAGCAGAGACCAAGCCCATGAGAGACTG 660
QY      884 CGGAGCTGGAGAGAGAGAGAGAGAGCCCGCTCAGATTTCAGAGGTGACAGCCGCTGGAG 943
Db      661 CGTGGCTGGAGAGAGAGAGAGAGAGCCCACTCAGGTTTCCAGGGTGAACAGCACTGGAG 720
QY      944 CAGTCTGGCTGTACCACTTGGGTGATGAGAGACATCGAGAGAGAGAGAGAGAGAGAGCTTA 1003
Db      721 CAGCCAGACTGTACCACTTGGGTGATGAGAGACATGAGAGAGAGAGAGAGAGAGAGAGAGCTTA 780

```


[illegible]

XX	Sequence	2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;
Query Match	42.1%; Score 783; DB 4; Length 2379;	
Best Local Similarity	100.0%; Pred. No. 1e-146;	
Matches 783; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	1046 GGCTCCCTTCCTCGTGGCCAGAAAGTCGAACCTGCCCCCGCACCTCCATCCCATCTGA	1105
Db	840 GGCTCCCTTCCTCGTGGCCAGAAAGTCGAACCTGCCCCCGCACCTCCATCCCATCTGA	899
OY	1106 TCTTGCTCCATGAGCGGGAAAGCCATCATCTCCCAACACGAAAGGCCAAGGCGTGAC	1165
Db	900 TCTTGCTCCATGAGCGGGAAAGCCATCATCTCCCAACACGAAAGGCCAAGGCGTGAC	959
OY	1166 CCGGCTCTCTTCCACTTCTCTTGAACCCCAATCCGCAAGGTCGTGAGCTCCAGCAAG	1225
Db	960 CCGGCTCTCTTCCACTTCTCTTGAACCCCAATCCGCAAGGTCGTGAGCTCCAGCAAG	1015
OY	1226 CTCCGGGGCAACCAAGAGGGAGCAAGCATTTTGTAGGTCTCCCAAGGCCCAAGGCAAG	1285
Db	1020 CTCCGGGGCAACCAAGAGGGAGCAAGCATTTTGTAGGTCTCCCAAGGCCCAAGGCAAG	1075
OY	1286 AGTGTGGGTGTGGGCCCATGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGACCCGCC	1345
Db	1080 AGTGTGGGTGTGGGCCCATGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGCCGCC	1133
OY	1346 ATCCCTGGAGTGTCCCCCTCCGCCACCTGTGCTGCGACCCCGGCTCTCCCTCCCTA	1405
Db	1140 ATCCCTGGAGTGTCCCCCTCCGCCACCTGTGCTGCGACCCCGGCTCTCCCTCCCTA	1195
OY	1406 GCCCCCTCTGGGGCAACAGAGCAACGATCCGAGCCAAAGAGAGTCACAGAGGCTGCCG	1465
Db	1200 GCCCCCTCTGGGGCAACAGAGCAACGATCCGAGCCAAAGAGAGTCACAGAGGCTGCCG	1255
OY	1466 GGCCCTGACGACCACTGGCTCAAGGTGGCCCTGTCTGGGGCGGAGCAACTGCGGGAG	1525
Db	1260 GGCCCTGACGACCACTGGCTCAAGGTGGCCCTGTCTGGGGCGGAGCAACTGCGGGAG	1315
OY	1526 CTGCCCGCTTGTGTGTGTATGAGACCGAGCCCGGGACGCCGCTCAAGACATGACAC	1585
Db	1320 CTGCCCGCTTGTGTGTGTATGAGACCGAGCCCGGGACGCCGCTCAAGACATGACAC	1375
OY	1586 CACCAACCACTGAACATCAACCATTAACACACTTCTACAGACATAGAGCCCTCC	1645
Db	1380 CACCAACCACTGAACATCAACCATTAACACACTTCTACAGACATAGAGCCCTCC	1433
OY	1646 CCAGGGCCCCCCTGCCATATGAAGACCCACCCCGGACACCAAGGCACTTATAT	1705
Db	1440 CCAGGGCCCCCCTGCCATATGAAGACCCACCCCGGACACCAAGGCACTTATAT	1495
OY	1706 CTATTAATTAATTTATTAATGATGATTAATTTGTATTAATTAATTAATTTACCCCATAT	1765
Db	1500 CTATTAATTAATTTATTAATGATGATTAATTTGTATTAATTAATTAATTTACCCCATAT	1555
OY	1766 ATTATGCTAGGCTCATGTAGAAAGTCATGTGAAACACAGAACTAACTTTATTTATAT	1825
Db	1560 ATTATGCTAGGCTCATGTAGAAAGTCATGTGAAACACAGAACTAACTTTATTTATAT	1615
OY	1826 GTT 1828	
Db	1620 GTT 1622	
RESULT 7		
AL003216		
ID	AL003216 standard; DNA; 2379 BP.	
XX		
AC	AL003216;	
XX		
DT	21-NOV-2001 (first entry)	
XX		
DE	Human reproductive system related antigen DNA SEQ ID NO: 5904.	

XX	Human, reproductive system related antigen; reproductive system disorder;	PR	21-SEP-2000;	2000US-0234274P
KM	Cancer; gene therapy; ds.	PR	25-SEP-2000;	2000US-0234998P
XX		PR	26-SEP-2000;	2000US-0235464P
XX	Homo sapiens.	PR	27-SEP-2000;	2000US-0235834P
OS		PR	27-SEP-2000;	2000US-0235836P
XX		PR	29-SEP-2000;	2000US-0236337P
PN	WO200155320-A2.	PR	29-SEP-2000;	2000US-0236367P
XX		PR	29-SEP-2000;	2000US-0236368P
PD		PR	29-SEP-2000;	2000US-0236369P
XX		PR	29-SEP-2000;	2000US-0236370P
PF		PR	02-OCT-2000;	2000US-0237037P
XX		PR	02-OCT-2000;	2000US-0237038P
XX		PR	02-OCT-2000;	2000US-0237039P
XX		PR	02-OCT-2000;	2000US-0237040P
XX		PR	13-OCT-2000;	2000US-0239935P
XX		PR	13-OCT-2000;	2000US-0239937P
XX		PR	13-OCT-2000;	2000US-0240960P
XX		PR	20-OCT-2000;	2000US-0241221P
XX		PR	20-OCT-2000;	2000US-0241785P
XX		PR	20-OCT-2000;	2000US-0241786P
XX		PR	20-OCT-2000;	2000US-0241787P
XX		PR	20-OCT-2000;	2000US-0241808P
XX		PR	20-OCT-2000;	2000US-0241809P
XX		PR	20-OCT-2000;	2000US-0241826P
XX		PR	01-NOV-2000;	2000US-0246417P
XX		PR	08-NOV-2000;	2000US-0246474P
XX		PR	08-NOV-2000;	2000US-0246475P
XX		PR	08-NOV-2000;	2000US-0246476P
XX		PR	08-NOV-2000;	2000US-0246477P
XX		PR	08-NOV-2000;	2000US-0246478P
XX		PR	08-NOV-2000;	2000US-0246523P
XX		PR	08-NOV-2000;	2000US-0246524P
XX		PR	08-NOV-2000;	2000US-0246525P
XX		PR	08-NOV-2000;	2000US-0246526P
XX		PR	08-NOV-2000;	2000US-0246527P
XX		PR	08-NOV-2000;	2000US-0246528P
XX		PR	08-NOV-2000;	2000US-0246532P
XX		PR	08-NOV-2000;	2000US-0246609P
XX		PR	08-NOV-2000;	2000US-0246610P
XX		PR	08-NOV-2000;	2000US-0246611P
XX		PR	08-NOV-2000;	2000US-0246613P
XX		PR	17-NOV-2000;	2000US-0249207P
XX		PR	17-NOV-2000;	2000US-0249208P
XX		PR	17-NOV-2000;	2000US-0249210P
XX		PR	17-NOV-2000;	2000US-0249210P
XX		PR	17-NOV-2000;	2000US-0249211P
XX		PR	17-NOV-2000;	2000US-0249212P
XX		PR	17-NOV-2000;	2000US-0249213P
XX		PR	17-NOV-2000;	2000US-0249215P
XX		PR	17-NOV-2000;	2000US-0249215P
XX		PR	17-NOV-2000;	2000US-0249216P
XX		PR	17-NOV-2000;	2000US-0249217P
XX		PR	17-NOV-2000;	2000US-0249218P
XX		PR	17-NOV-2000;	2000US-0249244P
XX		PR	17-NOV-2000;	2000US-0249245P
XX		PR	17-NOV-2000;	2000US-0249246P
XX		PR	17-NOV-2000;	2000US-0249265P
XX		PR	17-NOV-2000;	2000US-0249297P
XX		PR	17-NOV-2000;	2000US-0249299P
XX		PR	17-NOV-2000;	2000US-0249300P
XX		PR	01-DEC-2000;	2000US-0250316P
XX		PR	01-DEC-2000;	2000US-0250317P
XX		PR	05-DEC-2000;	2000US-0251030P
XX		PR	05-DEC-2000;	2000US-0251968P
XX		PR	05-DEC-2000;	2000US-0251799P
XX		PR	06-DEC-2000;	2000US-0251856P
XX		PR	08-DEC-2000;	2000US-0251868P
XX		PR	08-DEC-2000;	2000US-0251869P
XX		PR	08-DEC-2000;	2000US-0251969P
XX		PR	08-DEC-2000;	2000US-0251990P

PR	1-SEP-2000	2000US-0234274P
PR	25-SEP-2000	2000US-0234997P
PR	25-SEP-2000	2000US-0234998P
PR	26-SEP-2000	2000US-0235464P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235836P
PR	29-SEP-2000	2000US-0236337P
PR	29-SEP-2000	2000US-0236367P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236359P
PR	29-SEP-2000	2000US-0236370P
PR	29-SEP-2000	2000US-0236803P
PR	29-SEP-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	02-OCT-2000	2000US-0237040P
PR	13-OCT-2000	2000US-0239935P
PR	13-OCT-2000	2000US-0239937P
PR	13-OCT-2000	2000US-0239937P
PR	20-OCT-2000	2000US-0241123P
PR	20-OCT-2000	2000US-0241178P
PR	20-OCT-2000	2000US-0241786P
PR	20-OCT-2000	2000US-0241787P
PR	20-OCT-2000	2000US-0241808P
PR	20-OCT-2000	2000US-0241809P
PR	20-OCT-2000	2000US-0241826P
PR	01-NOV-2000	2000US-0246417P
PR	01-NOV-2000	2000US-0246474P
PR	08-NOV-2000	2000US-0246475P
PR	08-NOV-2000	2000US-0246476P
PR	08-NOV-2000	2000US-0246477P
PR	08-NOV-2000	2000US-0246478P
PR	08-NOV-2000	2000US-0246553P
PR	08-NOV-2000	2000US-0246553P
PR	08-NOV-2000	2000US-0246554P
PR	08-NOV-2000	2000US-0246555P
PR	08-NOV-2000	2000US-0246556P
PR	08-NOV-2000	2000US-0246611P
PR	08-NOV-2000	2000US-0246613P
PR	17-NOV-2000	2000US-0249207P
PR	17-NOV-2000	2000US-0249208P
PR	17-NOV-2000	2000US-0249209P
PR	17-NOV-2000	2000US-0249210P
PR	17-NOV-2000	2000US-0249211P
PR	17-NOV-2000	2000US-0249212P
PR	17-NOV-2000	2000US-0249213P
PR	17-NOV-2000	2000US-0249214P
PR	17-NOV-2000	2000US-0249215P
PR	17-NOV-2000	2000US-0249216P
PR	17-NOV-2000	2000US-0249217P
PR	17-NOV-2000	2000US-0249218P
PR	17-NOV-2000	2000US-0249245P
PR	17-NOV-2000	2000US-0249245P
PR	17-NOV-2000	2000US-0249246P
PR	17-NOV-2000	2000US-0249257P
PR	17-NOV-2000	2000US-0249259P
PR	17-NOV-2000	2000US-0249300P
PR	17-NOV-2000	2000US-0249301P
PR	01-DEC-2000	2000US-0250315P
PR	01-DEC-2000	2000US-0250316P
PR	05-DEC-2000	2000US-0251030P
PR	05-DEC-2000	2000US-0251988P
PR	05-DEC-2000	2000US-0256719P
PR	06-DEC-2000	2000US-0257147P
PR	08-DEC-2000	2000US-0251865P
PR	08-DEC-2000	2000US-0251866P
PR	08-DEC-2000	2000US-0251869P
PR	08-DEC-2000	2000US-0251969P
PR	08-DEC-2000	2000US-0251990P

XX 30-JUL-2001; 2001US-00918995.
PF
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
PI WPI; 2003-615964/58.
DR
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 8686; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030073623
XX
XX Sequence 427 BP; 84 A; 177 C; 112 G; 53 T; 0 U; 1 Other;
SQ
Query Match 22.5%; Score 418.2; DB 9; Length 427;
Best Local Similarity 99.3%; Pred. No. 5.6e-74;
Matches 420; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1076 CTGCCCCCGCACCTCCATCCCATCTGATCTGCTCCATGAGCCGAAAGCCATCCAC 1135
Db 4 CTGCCCCCGCACCTCCATCCCATCTGATCTGCTCCATGAGCCGAAAGCCATCCAC 63
QY 1136 ATCCCAACCGAAAGCCCGAAGCGGTGAGACCGGCGCTCTTCCACTTCTTGACACCCCA 1195
Db 64 ATCCCAACCGAAAGCCCGAAGCGGTGAGACCGGCGCTCTTCCACTTCTTGACACCCCA 123
QY 1196 ATGCGCAAGGCTCAGAGCTCCAGCAACGCTCGGGGCAACCGGAGCGGGAAGAGAC 1255
Db 124 ATGCGCAAGGCTCAGAGCTCCAGCAACGCTCGGGGCAACCGGAGCGGGAAGAGAC 183
QY 1256 TTGTGAGGTCTCCCAAGGCGCCAGGAGCAAGATGTGGGTGTGGGCCAAGTGGCCAGAGGG 1315
Db 184 TTGTGAGGTCTCCCAAGGCGCCAGGAGCAAGATGTGGGTGTGGGCCAAGTGGCCAGAGGG 243
QY 1316 GCAAGAAACAAAGCCCTCTTGGAGACCGGCAATCCTTGCGGTGTCTCCCTCGCCACCTG 1375
Db 244 GCAAGAAACAAAGCCCTCTTGGAGACCGGCAATCCTTGCGGTGTCTCCCTCGCCACCTG 303
QY 1376 GGTGCGAGCGCGGCGCTCTCCCTCTAGCGCCCGCGGAGCAAGAAAGCAGAGAC 1435
Db 304 GGTGCGAGCGCGGCGCTCTCCCTCTAGCGCCCGCGGAGCAAGAAAGCAGAGAC 363
QY 1436 CGAGCCAAAGAGAGACGACGAGAGGCTGCGGGGCTTGAGAGCACCAGTGGCTCTCAGTGGC 1495
Db 364 CGAGCCAAAGAGAGACGACGAGAGGCTGCGGGGCTTGAGAGCACCAGTGGCTCTCAGTGGC 423

QY 1496 CCT 1498
Db 424 CCT 426
RESULT 11
ID AAS78752 standard; cDNA; 1743 BP.
XX
XX AAS78752;
AC
XX 13-FEB-2002 (first entry)
DT
XX
DE DNA encoding novel human diagnostic protein #14556.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200175067-A2.
PN
XX 11-OCT-2001.
PD
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR P-PSDB; ABG14565.
DR
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PT
XX
XX Claim 1; SEQ ID NO 14556; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;
Query Match 15.3%; Score 284.2; DB 5; Length 1743;
Best Local Similarity 77.8%; Pred. No. 5.4e-47;
Matches 410; Conservative 0; Mismatches 3; Indels 114; Gaps 1;
QY 1047 GCTCCCTTCGTGAGCCCAAGTCAAGACTGCCCCCGGAGCTCCATCCACTGAT 1106

```
Db 332 GCTCCCTTCCTGCGCCCAAGTCAAGCTGCCCCCGCACTCCAAATCCACTGAT 391
Qy 1107 CTGCTCCCATGAGCGGGAAGCCATCCATCCACACCGAAAGCCCAAGCGTGAGCC 1166
Db 392 CTGCTCCCATGAGCGGGAAGCCATCCATCCACACCGAAAGCCCAAGCGTGAGCC 451
Qy 1167 CGGCTCTCTTCCACTTCTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGACGGC 1226
Db 452 CGGCTCTCTTCCACTTCTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGACGGC 511
Qy 1227 TCCGGGGGCAACCGAGAGGAGGACCTTGTGAGTCCCGCCCAAGGCGCCAGAG 1286
Db 512 TCCGGGGGCAACCGAGAGGAGGACCTTGTGAGGCGCCCGCCCAAGGCGCAAAA 571
Qy 1287 GTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAGCCCTCTGGGACCCGCA 1346
Db 572 GTGTG-----G 576
Qy 1347 TCCCTGGGTGTCCCTCCCGCCCACTGGCTGCCAGCCCGGCGCTCTCCCTCTAG 1406
Db 577 -----G 577
Qy 1407 CCCCCCTCGGGCAAGAGACACAGACCGAGGCGAAGAGGACGAGGCGTGCCGG 1466
Db 578 CCCCCCTCGGGCAAGAGACACAGACCGAGGCGAAGAGGACGAGGCGTGCCGG 637
Qy 1467 GCCTGACGACCACTGCGCTCAGGTGGCCCTGTCTGGGCGGAGACACTGCGGAGC 1526
Db 638 GCCTGACGACCACTGCGCTCAGGTGGCCCTGTCTGGGCGGAGACACTGCGGAGC 697
Qy 1527 TGCCCGCTTGTGTGTATGAGAGCCAGGCGGCGAGCCGCTTCAG 1573
Db 698 TGCCCGCTTGTGTGTATGAGAGCCAGGCGGCGAGCCGCGCCAG 744
```

RESULT 12

AAS92081 standard; cDNA; 1743 BP.

AAS92081;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #27885.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PsDB; ABG27894.

New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.

Claim 1; SEQ ID NO 27885; 103bp; English.

```
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
```

Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Query Match 15.3%; Score 284.2; DB 5; Length 1743;

Best Local Similarity 77.8%; Pred. No. 5,4e-47;

Matches 410; Conservative 0; Mismatches 3; Indels 114; Gaps 1;

```
Qy 1047 GCTCCCTTCCTGCGCCCAAGTCAAGCTGCCCCCGCACTCCAAATCCACTGAT 1106
Db 332 GCTCCCTTCCTGCGCCCAAGTCAAGCTGCCCCCGCACTCCAAATCCACTGAT 391
Qy 1107 CTGCTCCCATGAGCGGGAAGCCATCCATCCACACCGAAAGCCCAAGCGTGAGCC 1166
Db 392 CTGCTCCCATGAGCGGGAAGCCATCCATCCACACCGAAAGCCCAAGCGTGAGCC 451
Qy 1167 CGGCTCTCTTCCACTTCTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGACGGC 1226
Db 452 CGGCTCTCTTCCACTTCTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGACGGC 511
Qy 1227 TCCGGGGGCAACCGAGAGGAGGACCTTGTGAGTCCCGCCCAAGGCGCCAGAG 1286
Db 512 TCCGGGGGCAACCGAGAGGAGGACCTTGTGAGGCGCCCGCCCAAGGCGCAAAA 571
Qy 1287 GTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAGCCCTCTGGGACCCGCA 1346
Db 572 GTGTG-----G 576
Qy 1347 TCCCTGGGTGTCCCTCCCGCCCACTGGCTGCCAGCCCGGCGCTCTCCCTCTAG 1406
Db 577 -----G 577
Qy 1407 CCCCCCTCGGGCAAGAGACACAGACCGAGGCGAAGAGGACGAGGCGTGCCGG 1466
Db 578 CCCCCCTCGGGCAAGAGACACAGACCGAGGCGAAGAGGACGAGGCGTGCCGG 637
Qy 1467 GCCTGACGACCACTGCGCTCAGGTGGCCCTGTCTGGGCGGAGACACTGCGGAGC 1526
Db 638 GCCTGACGACCACTGCGCTCAGGTGGCCCTGTCTGGGCGGAGACACTGCGGAGC 697
Qy 1527 TGCCCGCTTGTGTGTATGAGAGCCAGGCGGCGAGCCGCTTCAG 1573
Db 698 TGCCCGCTTGTGTGTATGAGAGCCAGGCGGCGAGCCGCGCCAG 744
```

RESULT 13

ADC31966 standard; cDNA; 1743 BP.

ADC31966;

18-DEC-2003 (first entry)

XX Human novel cDNA contig sequence, SEQ ID NO:2048.
DE
XX
XX Human, diagnostic; drug screening; forensics; gene mapping;
KM biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KM neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KM ulcer; osteoporosis; autoimmune disease; cancer;
KM molecular weight marker; food supplement; antiparkinsonian; nootropic;
KM neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KM anticancer; osteopathic; immunosuppressive; antiinflammatory; cyostatic;
KM gene therapy; ss.
XX
XX Homo sapiens.
OS
XX W02003029271-A2.
PN
XX 10-APR-2003.
PD
XX 24-SEP-2002; 2002WO-US030474.
PF
XX 24-SEP-2001; 2001US-0324631P.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Dzmanac RT;
XX
XX WPI; 2003-371981/35.
DR P-PSDB; ADC32733.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemia, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
XX Example 2; SEQ ID NO 2048; 1185bp; English.
PS
XX The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC coding sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune disease or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Query Match 15.3%; Score 284.2; DB 10; Length 1743;
Best Local Similarity 77.8%; Pred. No. 5.4e-47;
Matches 410; Conservative 0; Mismatches 3; Indels 114; Gaps 1;
QY 1047 GCTCCCTTCCTGAGCCAGAGTCAAGTCTGCCCCCGACCTCCATCCACTGAT 1106
DB 332 GCTCCCTTCCTGAGCCAGAGTCAAGTCTGCCCCCGACCTCCATCCACTGAT 391
QY 1107 CTGCTCCCATGAGCCGGAAGCCATCCATCCCAACGGAAGCCCAAGGCGTGAAC 1166
DB 392 CTGCTCCCATGAGCCGGAAGCCATCCATCCCAACGGAAGCCCAAGGCGTGAAC 451
QY 1167 CGGCTTCCTTCACCTTCCTTGAACACCCCAATCGCAAGTCTGAGCTCAGCAAGGC 1226
DB 452 CGGCTTCCTTCACCTTCCTTGAACACCCCAATCGCAAGTCTGAGCTCAGCAAGGC 511
QY 1227 TCCGGGACACCCAGAGCAGGAGCAAGCACTTTGTAGTCCCCCAAGGCCAAGAGA 1286
DB 512 TCCGGGACACCCAGAGCAGGAGCAAGCACTTTGTAGTCCCCCAAGGCCAAGAGA 571
QY 1287 GTGTGGGTGTGGGCGACGTGGCCAGAGGGGCAAGAAACAGCCCTCTGGACCCGCCA 1346
DB 572 GTGTG----- 576
QY 1347 TCCCTGCGTGTCCCTCCGCGCACCTGAGTGCAGCCGCGCTCTCCCTCCTAG 1406
DB 577 -----G 577
QY 1407 CCCCCCTCGGACCAAGAGCACAGAGCAGGCAAGAGAGCAGAGGCTGCGGG 1466
DB 578 CCCCCCTCGGACCAAGAGCACAGAGCAGGCAAGAGAGCAGAGGCTGCGGG 637
QY 1467 GCTGACAGGACCACTGCGCTCAGGTGGCCCTGTCTCTGGGGCGGAGACCTGCGGAGC 1526
DB 638 GCTGACAGGACCACTGCGCTCAGGTGGCCCTGTCTCTGGGGCGGAGACCTGCGGAGC 697
QY 1527 TGCCCGCTGTGTGTATGAGAGCCAGGCGGAGCGGTCCAG 1573
DB 698 TGCCCGCTGTGTGTATGAGAGCCAGGCGGAGCGGCGCCAG 744
RESULT 14
AAS27117
ID AAS27117 standard; cDNA; 1966 BP.
AC AAS27117;
XX
DT 07-NOV-2001 (first entry)
XX
XX cDNA encoding novel signal transduction pathway protein, Seq ID 152.
DE
XX Neuroprotective; cyostatic; dermatological; immunosuppressive; tumour;
KM antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KM immune system disorder; rheumatoid arthritis; inflammatory condition;
KM organ transplant rejection; infection; hepatitis C; blood disorder;
KM sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KM neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KM chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KM cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KM reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KM acquired immune deficiency syndrome.
XX
XX Homo sapiens.
OS
XX W0200154733-A1.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001312.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235464P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244517P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250381P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
XX
PA Rosen CA, Barash SC, Ruben SM;
XX
PI WPI; 2001-465460/50.
XX
DR P-PSDB; AAU17200.
DR
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
XX diagnosing disorders related to the proteins, including cancers, immune
PS disorders and neuronal disorders.
PS Claim 1; SEQ ID NO 152; 880bp; English.

XX The invention relates to novel isolated polypeptides (I), and
CC polymucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haemotopoietic disorders, hyperproliferative disorders
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC respiratory disorders, dermatological disorders, in wound healing,
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders, gastrointestinal disorder
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
CC B-cell responsiveness to pathogens, activators of T-cells, to induce
CC higher affinity antibodies, and as a means to induce tumour proliferation
CC in pathology e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
CC AAS27850 represent novel signal transduction pathway protein coding
CC sequences and PCR primers of the invention
XX

Query Match	14.7%	Score 272.4;	DB 4;	Length 1966;
Best Local Similarity	56.5%	Pred. No. 1.3e-44;		
Matches 667; Conservative	0;	Mismatches 471;	Indels 42;	Gaps 7;

OY	497	GAGAAAGATGACCGGCTGTGGGACCGGAGATATGAAAGAAAGATGAGAGAGTGAAGCGAACC	556
Db	516	GAGCAACCCGGGACAACCTCTTCAAGCCCAATATGACGAGAGGGGACGAAACCGGCGAAGGGC	575
OY	557	TGCCAGGCTC---CAAGAGCAGCTGAAGTTTGAAGAGCTTCAAGTCGAGCTGTGTCATG	613
Db	576	CCGGAGGAGACCGGGGCGGCGAGGGCTCAACATTTGACGACTTCAGTGGAGATGTCTCGGTG	635
OY	614	GAGGAGGACAGCCCGGCGAGAGTGGACCTTCAACCTGTATGACTTTGACAAACAGCGCAAG	673
Db	636	GAGGAGGACGACCCGCGAAGATGACGCTTCAAGCTCTATGACTTTGACAACTGCGGGAAAG	695
OY	674	GTCACCCGAGAGGACATCACCAGCTGTGTGGACACATCTATGAGGTGTGGAATCTCCCT	733
Db	696	GTCACAGGAGAGGACATGTCCAGCTTCAATGACACACATCTTGAAGGTGTGGAATGCTTCG	755
OY	734	GTCAAACACTCCCAACATCCAGCAGAATGCTGCGGGTAAAGCTCACCGTGGCCCCGAT	793
Db	756	GTCAAACACTCTCGGGGACGAGCAAGACCCTCGGTGAAGCTAACCGTCAGCCCTGAG	815
OY	794	GGCAGCCAGAGCAAGAGAGAGTCTTGTCAATCAGGCTGACCTTGCAGAGGCGACAGGCC	853
Db	816	CCCTTCAGCAAGAGAGAGAGGTCTCTCTGTGGCCAGAGACCGGAGACCCACCCGTTGGC	875
OY	854	CGAGCAGAGACCAAGCCCACTGAGGACCTGCGGAGCTGGAGAAAGAACAGCGACCCCG	913
Db	876	AGGATGAGAGGGTGAACCTGGCAGAGGAGCCAAAGGTGGCTGACAGGAGGTTGTCTGCACAC	935
OY	914	CTCAGGTTCCAGGGTGAAGCAGCCGCGCTGAGCAGTCTGGCTCTACACCATTTGCGTAGAT	973
Db	936	GTCAGGAGGCCAGTACTGACCCCCAGCCCTGCTGGAGCGGGGGCCCTTACTGCGTGGAC	995
OY	974	GAGAACATTCGAGAGAGAAACCACTACTTAATCTCTGCGCGGGATTTGAAAATTAACAGCTCC	1033
Db	996	GAGAAACAGGAGCCGAGAAACCACTACTTGGACCTCTGCGGGGATTTGAAGACTTACACGTCC	1055
OY	1034	CAATTTGGGCTTGAGCTCCCTTTCCTGTGACCCAGAAAGTCAGAACTGCCCCCGCACCTCC	1093
Db	1056	AGATTTCGGGCTCTGGTCTCCCTCTCTGTGCAAGCAAGACAGAGGCCCCAGAGGAGGGCTCTCG	1115
OY	1094	AATCCCA---CTCGATCTTGCTCTCCCATGAGCCGGAAGCATTCACATCTCCACACCGAAG	1150
Db	1116	CACCTTCAGGCCCCGGTCTCCGCTCCACAGAGCCAGTATACATGCGGTATACACACCGCAGG	1175

QY	1151	CCCCAAGGCGTGGACCCGGGCGTCTTCCATCTTCCATGACACCCCAATCGCCAAAGTCTCA	1210
Db	1176	TCACAGGTCGTGTGAAACACGTGTGTGCAGCTTCGAH--GCTTGTGCTCCGGGCGCTTG	1232
QY	1211	GAGCTCCAGCAACGCTCTCCGGGGACCAACAGACGGAGACAATTTGTAGGTCTCCC	1270
Db	1233	GACACGAGAGCCCCGGCGGAAGGGGCGGAG-----AAGCAGTTCTCTCAATCTCCCC	1283
QY	1271	AAGGCCAAGGGCAAGAGTGTGGTGTGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCC	1330
Db	1284	AAGGCTCTCCGGAAGCCGCGCTGGGGGTGCAGCCACAGACAAGTCCGGGAAGCTTTAGC	1343
QY	1331	CCTCTGGGAGCCCGGCATCTCCCTGCGGATGCCCTCCGCGCCACCTGAGCTGCAGCCGAGC	1390
Db	1344	TACTACGTGCGCGCGCTC-----TGCGGCGCCAGGCGCCTCTAGGACGGCCACCACTTC	1397
QY	1391	CTCTCTCCCTCTCTTAGCCCCCTTGGGGCAAGAAGCAACAGCACCGAGCTCAGAGAGC	1450
Db	1398	CCGACAGCCCCCGACCGCACCTTAAGGGCCACAAGAGGGTATCCGCGCAAAAGGGGAGGAAGGC	1457
QY	1451	CAGCAGGCGTGCAGGGGCGTGCAGAGCAACACTGCGCTCAGAGTGAGCCTGTCTGAGGCGG	1510
Db	1458	CAGTGGCACTCAAGGCCCCACAGAGCTCAGCTTCCACAG-----TGAGACAC	1505
QY	1511	GAGCACTCGCGGAGCTGCCCGCGCTTGATGTATAGAGCCAGGCGGGAGCCGCTC	1570
Db	1506	GAGGTGTGTGCGGAGCTTGCAGCGCCAGCAGAGAGAGGAGCTAAGCGGTGCCAGTGATC	1565
QY	1571	CAGAGACATGAGACACACACACACATGAAATCATACACCAATT-----ACACACACTTC	1624
Db	1566	CAGCGGCACGAGCACACACACACACAGAGACACACACACACACACACACACCACTTC	1625
QY	1625	TACCAAGCATAGAGACCCCTCCCGAGGGCGCCACCCCTGCACA	1664
Db	1626	CACCGTCTTAGCGCCACTGCCAAGACACCTCGCTCCCA	1665
RESULT 15			
ADB93295			
ID	ADB93295	standard; cDNA; 1966 BP.	
XX	AC	ADB93295;	
XX	DT	04-DEC-2003 (first entry)	
XX	DE	Human cDNA encoding a novel protein #142.	
KW	XX	as; gene; human; autoimmune disease; Parkinson's disease; silicosis;	
KW	XX	gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;	
KW	XX	immunosuppressive agent; adjuvant; enhance immune response;	
KW	XX	higher affinity antibody induction;	
XX	OS	increased serum immunoglobulin concentration.	
XX	XX	Homo sapiens.	
XX	XX	US2002168711-A1.	
XX	PD	14-NOV-2002.	
XX	PF	17-JAN-2001; 2001US-00764868.	
XX	PR	31-JAN-2000; 2000US-0179065P.	
XX	PR	04-FEB-2000; 2000US-0180628P.	
XX	PR	28-JUN-2000; 2000US-0214886P.	
XX	PR	07-JUL-2000; 2000US-0216647P.	
XX	PR	07-JUL-2000; 2000US-0216880P.	
XX	PR	11-JUL-2000; 2000US-0217496P.	
XX	PR	11-JUL-2000; 2000US-0217496P.	
XX	PR	14-JUL-2000; 2000US-0218290P.	
XX	PR	26-JUL-2000; 2000US-0220963P.	
XX	PR	26-JUL-2000; 2000US-0220964P.	
XX	PR	26-JUL-2000; 2000US-0220964P.	
XX	PR	14-AUG-2000; 2000US-0224518P.	
XX	PR	14-AUG-2000; 2000US-0224519P.	

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226688P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 17-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.

PI Rosen CA, Ruben SM, Barash SC;

WPI; 2003-719985/68.

DR P-PSDB; ADB93908.

PT New isolated polypeptide useful for diagnosing and treating
PT Immunosuppressive conditions such as autoimmune disease and Parkinson's
XX disease.

PS Claim 3; SEQ ID NO 152; 345dp; English.

XX The invention relates to an isolated polypeptide. The polypeptide is
CC useful for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, by determining the presence or
CC amount of expression of the polypeptide in a biological sample and
CC diagnosing a pathological condition or a susceptibility to a pathological
CC condition based on the presence or amount of expression of the
CC polypeptide. The polypeptide is also useful for identifying a binding
CC partner to the polypeptide, which involves contacting the polypeptide
CC with a binding partner and determining whether the binding partner
CC effects an activity of the polypeptide. The polypeptide or the nucleic
CC acid encoding the polypeptide is useful for preventing, treating, or
CC ameliorating a medical condition, which involves administering the
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
CC is useful for diagnosing a pathological condition or a susceptibility to
CC a pathological condition in a subject, which involves determining the
CC presence or absence of a mutation in the nucleic acid, and diagnosing a
CC pathological condition or susceptibility to a pathological condition
CC based on the presence or absence of the mutation. The polypeptide, the
CC nucleic acid and an antibody to the polypeptide are useful for treating

CC autoimmune disease, Parkinson's disease, siltosis, gastrointestinal
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
CC the nucleic acid and the antibody are useful as immunosuppressive agents,
CC as adjuvants to enhance immune responses, and as agents to induce higher
CC affinity antibodies and increase serum immunoglobulin concentrations. The
CC present sequence represents cDNA encoding a novel human protein. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic form direct from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20020168711.

XX Sequence 1966 BP; 391 A; 726 C; 608 G; 241 T; 0 U; 0 Other;

Query Match 14.7%; Score 272.4; DB 10; Length 1966;

Best Local Similarity 56.5%; Pred. No. 1.3e-44;

Matches 667; Conservative 0; Mismatches 471; Indels 42; Gaps 7;

QY 497 GAGAACTGACGGGCTGGGACGGGAGAGAGAAAGATGAGAGAGTGGAGAACCC 556
DB 516 GAGACCCGGGAGACACTCTCTCAGGACGATGACGAGAGAGGAGCAACCCGGAGGGC 575
QY 557 TGCCAGGCTC---CAAGAGCAGCTGAAGTTGAAGAGCTCCAGTGCAGTGTCCATG 613
DB 576 CCGGAGGAGACGGGGGGGAGCGCCCTCAATTGACGACTCCAGTGCAGTGTCTGGTG 635
QY 614 GAGAGAGACAGCGGCGAGAGTGAACCTTCACTGTATGACTTTGACAAACAGCGCAG 673
DB 636 GAGAGAGACAGCGGCGAGAGTGAACCTTCACTGTATGACTTTGACAAACAGCGCAG 695
QY 674 GTCAACCGAGAGACATCAACAGCTTGTGACACACCTATAGAGGAGTGGACTCTCT 733
DB 696 GTCAACCGAGAGACATCAACAGCTTGTGACACACCTATAGAGGAGTGGACTCTCT 755
QY 734 GTCAACCACTCCCAACATCAACAGATCTCGGGTAAAGCTCAACCTGCCCCGAT 793
DB 756 GTCAACCACTCTCCGAGACAGCAAGACCTCCGTTGAAGCTTAACCGTCAAGCTTGAG 815
QY 794 GGCAGCCAGAGCAAGAGAGAGCTCTTGTCAATCAAGCTTCACTGACAGCGCAAGCCC 853
DB 816 CCTCCAGCAAGAGAGAGAGAGTCTCTGCTGAGCAGAGACCGGAGCCACCCGTTGC 875
QY 854 CGAGCAGAGCAAGAGCCACTGAGGAGCCTGCGAGCTGGAGAGAGAGAGAGAGAGCCCG 913
DB 876 AGATGAGAGGTGAATCTGCGAGAGAGAGCAAGAGGTGCTGACAGAGAGTGTCTGACAC 935
QY 914 CTCAGTTCCAGAGGTGACAGCCCTGAGCAGTGTGCTGACCACTTGGTGTGAT 973
DB 936 GTCAAGAGGCGCAGTACTGACCCCGAGCCCTGCTCGAGAGCGGGGCGCTTACTGTGGAC 995
QY 974 GAGAACATGAGAGAGAGAAACCACTTATGATCTGCGGGATGAGAAACTACAGTCC 1033
DB 996 GAGAACACGAGAGCGCAGAGAAACCACTTATGATCTGCGGGATGAGAAACTACAGTCC 1055
QY 1034 CAATTGGGGCTGGCTCCCTCCGAGGCCAGAGTCAAGATCGCCCCCGGACCTCC 1093
DB 1056 AGATTGGCTCTGAGTCCCTCTCTGTCAGAACAGAGAGAGAGCCCGGAGAGGCTTCG 1115
QY 1094 AATCCCA---CTGATCTCTGCTCCATGAGCCGGAAGCCATTCACATCCACAGCGAAG 1150
DB 1116 CACTTCAGGCGCGGTCCGCTCCAGAGCCAGATGACATGCGGTACACCGGACGAG 1175
QY 1151 CCCCAGGCGTGAACCGGCTCTCTTCACTTCTTGACACCCCAATCGCCAGAGTCTCA 1210
DB 1176 TCACAGGTGTGTGTAACAGTGTGTCACGCTTCGAA---GCCTGTGCGCGGCTCG 1232
QY 1211 GAGCTCAGCAAGAGGCTCGGGGACCCAGGAGCGGAGCAAGCACTTGTAGGTCCCGC 1270
DB 1233 GACACGACGCGCGCGCGGAGAGGAGCGGAGC-----AGGACTTCTCAAGTCCCGC 1283
QY 1271 AAGGCCAGGAGCAAGAGTGTGTGTGTGAGCCAGTGTGCGCAGAGAGGCGAGAGAAACAGCCC 1330
DB 1284 AAGGGGTCCGGAGAGCGGCTGTGGGTGCGCAGCAGAGCAAGTGTGCGGAGAAAGCTTCAAGC 1343
QY 1331 CTTCTGAGACCGGCAATCCTCTGCGGTGTCCCTCCGCGCCACCTGAGTGTGCAAGCCGCGC 1390

```

Db      1344 TACTACCTGCGGCGCTGC-----TGCGGCCAGGCCCCCTCAGGAGGCGCACCACTTC 1397
QY      1391 CTCTCTCCCTTCCTTAGCCCCCTTGCGCACAAGACACCGAGCCAGAGAGC 1450
Db      1398 CCGCAGCCCCCACCACCCCTACGCGCACAAAGCGTACCGCCAAAAGGCGAGGAGGCG 1457
QY      1451 CAGCAGGCGCTGCCGCGGCGCTGCAAGCACACTGGCTCAGGTGGCCCTGTCCTGGGCGG 1510
Db      1458 CACTGCCACTCAAGGCCCTCCACACGCTCAGCTGCACAG-----TGAGCAC 1505
QY      1511 GAGCACCCTGCGGAGCTGCCCCGCTTGCTGTATGAGACCAAGCCGCGGCGAGCCGATC 1570
Db      1506 GAGGTGGTGGGGAAGCTGCGCGCCACGCGCAGAGAGAGGCGTACGCGGTGCCAGTATC 1565
QY      1571 CAGAGACATAGACACCAACCAACCAATGAACATACCACTT-----ACCACTTTC 1624
Db      1566 CAGCGGACAGACACCAACCAACCAAGAGACCAACCAACCAACCAACCACTTC 1625
QY      1625 TACCAGCATAGAGCCCTCCCGCAGGGCCCCACCCCTGCCA 1664
Db      1626 CACCCTCTAGCGCCACTGCGCAAGCACACCTTCGCTCCA 1665

```

Search completed: December 29, 2004, 20:56:14
 Job time : 1312 secs

Db 61 CGGGAGTCGGGCGCGCGGCGAGAGAGCGGCTCCCGGCGCGCTCGGCGTCCG 120
Qy 121 CTGGGCTGGGCGCTGCTTGGGAGAGAGAGCGCAAGAGGCGCAAGCGCGCGCG 180
Db 121 CTGGGCTGGGCGCGCTGCTTGGGAGAGAGAGCGCAAGAGGCGCGCGCGCGCG 180
Qy 181 GGGGCGATGGGCTTAAAGGAGCGCTCCCGGCGCGGCGCGCGCGCGCGCGCGCG 240
Db 181 GGGGCGATGGGCTTAAAGGAGCGCTCCCGGCGCGGCGCGCGCGCGCGCGCGCG 240
Qy 241 CAAGCGCGCGCGCGCTGTCAGAGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCG 300
Db 241 CAAGCGCGCGCGCGCTGTCAGAGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCG 300
Qy 301 CGCTGCTGGGCTGTCAGAGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCG 360
Db 301 CGCTGCTGGGCTGTCAGAGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCG 360
Qy 361 TGTCTGGGAGCG 420
Db 361 TGTCTGGGAGCG 420
Qy 421 CTGGGCGAGCGTGTTCAGAGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCG 480
Db 421 CTGGGCGAGCGTGTTCAGAGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCG 480
Qy 481 AGTGGCGCTGCTCTCTGAGAGAGCTGACGGGCTGGGCGCGCGAGATGAGAGAG 540
Db 481 AGTGGCGCTGCTCTCTGAGAGAGCTGACGGGCTGGGCGCGCGAGATGAGAGAG 540
Qy 541 GAGAGTGAAGGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 541 GAGAGTGAAGGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Qy 601 CGAGCTGTCAATGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 601 CGAGCTGTCAATGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Qy 661 CAAGAGCGGCAAGGTCAACCGAGAGAGATCAAGCTTGGTCAACATCTATGAG 720
Db 661 CAAGAGCGGCAAGGTCAACCGAGAGAGATCAAGCTTGGTCAACATCTATGAG 720
Qy 721 GGTGGAATCTCTCTGTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 721 GGTGGAATCTCTCTGTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Qy 781 CGTGGCG 840
Db 781 CGTGGCG 840
Qy 841 GAGGCGCAAGGCGCGCGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db 841 GAGGCGCAAGGCGCGCGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 900
Qy 901 GCAGCGAGCGCGCGCTGAGGTTCAGAGGAGCGCGCGCGCGCGCGCGCGCG 960
Db 901 GCAGCGAGCGCGCGCTGAGGTTCAGAGGAGCGCGCGCGCGCGCGCGCGCG 960
Qy 961 CCATTTGGTATAGAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 CCATTTGGTATAGAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Qy 1021 AAATCTACGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1021 AAATCTACGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1081 CCGCGCGAGCTGCAATCCCATCTGATCTGCTCCCAAGAGCGAGAGAGAGAG 1140
Db 1081 CCGCGCGAGCTGCAATCCCATCTGATCTGCTCCCAAGAGCGAGAGAGAGAG 1140
Qy 1141 ACACCGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 ACACCGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

Qy 1201 CAAGTCTGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 CAAGTCTGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1261 GAGGTCCCGCAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 GAGGTCCCGCAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1321 AAACAAAGCGCGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 AAACAAAGCGCGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1381 CAGCGCGCGCGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 CAGCGCGCGCGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1441 CAAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 CAAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Qy 1501 CTGGGCGGAG 1560
Db 1501 CTGGGCGGAG 1560
Qy 1561 GCAGCGGCTGAG 1620
Db 1561 GCAGCGGCTGAG 1620
Qy 1621 CTCTACAG 1680
Db 1621 CTCTACAG 1680
Qy 1681 CCGGACAG 1740
Db 1681 CCGGACAG 1740
Qy 1741 TAAATTAATTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 1741 TAAATTAATTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 1801 CACGAG 1859
Db 1801 CACGAG 1859

RESULT 2
US-09-993-966-1
; Sequence 1, Application US/0993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NCD PROTEIN,
; FILE REFERENCE: 014024/028073
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-1

Query Match 96.1%; Score 1786; DB 10; Length 1786;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 AGTCGGGCGCGGCGAGGCGGCGTCCGCGCGCGCTCGGCTCCG 125
Db 1 AGTCGGGCGCGGCGAGGCGGCGTCCGCGCGCGCTCGGCTCCG 60

QY 126 CTGGGGGGCTGCTTCGGGAGAGAGAGCCAGAGAGAGCGCGCGCG 185
Db 61 CTGGGGGGCTGCTTCGGGAGAGAGAGCCAGAGAGAGCGCGCGCG 120

QY 186 CATGGCTTAAAGAGAGCTCCGCGCGCGCGAGCCAGAGAGAGAG 245
Db 121 CATGGCTTAAAGAGAGCTCCGCGCGCGCGAGCCAGAGAGAG 180

QY 246 CGGCGCGCGCTGTGCGAGCGAGAGAGCGCGAGAGAGAGAG 305
Db 181 CGGCGCGCGCTGTGCGAGCGAGAGAGCGCGAGAGAGAGAG 240

QY 306 CTTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
Db 241 CTTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

QY 366 CGGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
Db 301 CGGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

QY 426 GCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
Db 361 GCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

QY 486 CCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
Db 421 CCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480

QY 546 TGAAGCAACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 605
Db 481 TGAAGCAACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 540

QY 606 TGTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
Db 541 TGTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

QY 666 ACGGCAAGGTCACCCGAGAGAGAGAGAGAGAGAGAGAGAG 725
Db 601 ACGGCAAGGTCACCCGAGAGAGAGAGAGAGAGAGAGAGAG 660

QY 726 ACTGCTCTGTCAACCACTCCGCAACATCAAGAGAGAGAGAG 785
Db 661 ACTGCTCTGTCAACCACTCCGCAACATCAAGAGAGAGAGAG 720

QY 786 CCCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 845
Db 721 CCCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

QY 846 CAAGGCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905
Db 781 CAAGGCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840

QY 906 GAGGCGCGCTCAGGTTCCAGGAGAGAGAGAGAGAGAGAGAG 965
Db 841 GAGGCGCGCTCAGGTTCCAGGAGAGAGAGAGAGAGAGAGAG 900

QY 966 GCGTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
Db 901 GCGTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

QY 1026 ACAGCTCCCAATTTGGGCTCGGCTCCGCTCGGCGCGAGAG 1085
Db 961 ACAGCTCCCAATTTGGGCTCGGCTCCGCTCGGCGCGAGAG 1020

QY 1086 GCAAGCTCCCAATTTGGGCTCGGCTCCGCTCGGCGCGAGAG 1145
Db 1021 GCAAGCTCCCAATTTGGGCTCGGCTCCGCTCGGCGCGAGAG 1080

QY 1146 GAAAGCCCCAAGGCGTGGAGAGAGAGAGAGAGAGAGAGAG 1205
Db 1081 GAAAGCCCCAAGGCGTGGAGAGAGAGAGAGAGAGAGAGAGAG 1140

QY 1206 TCTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265
Db 1141 TCTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

QY 1266 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
Db 1201 CCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1326 AGCCCTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385
Db 1261 AGCCCTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

QY 1386 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
Db 1321 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380

QY 1446 AGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
Db 1381 AGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

QY 1506 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565
Db 1441 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

QY 1566 CGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
Db 1501 CGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560

QY 1626 ACCAGACATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685
Db 1561 ACCAGACATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620

QY 1686 CACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1745
Db 1621 CACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680

QY 1746 ATTATTTGTTACTCAGTAATTTAGTACCTTACATGTAAGAG 1805
Db 1681 ATTATTTGTTACTCAGTAATTTAGTACCTTACATGTAAGAG 1740

QY 1806 AACTAACTTTTATTTATTTATTTATTTATTTATTTATTTAT 1851
Db 1741 AACTAACTTTTATTTATTTATTTATTTATTTATTTATTTAT 1786

RESULT 3
US-10-087-192-179
Sequence 179, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
PRIOR FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 179
LENGTH: 1448
TYPE: DNA
ORGANISM: Homo sapiens
US-10-087-192-179

Query Match 77.8%; Score 1446.4; DB 13; Length 1448;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 186 CATGGCTTAAGGAGAGCTCCCGGCGCGGAGCCCGCAGCATGGGAAACTTCACTCCAAAGC 245
Db 1 CATGGCTTAAGGAGAGCTCCCGGCGCGGAGCATCCCAAGCATGGGAAACTTCACTCCAAAGC 60

QY 246 CGGCGCGCTGTGTGCAAGCGGAGGAGAGCCCGGAAAGGTGACAGTTGCGCGGTGAGCCGTG 305
Db 61 CGGCGCGCTGTGTGCAAGCGGAGGAGAGCCCGGAAAGGTGACAGTTGCGCGGTGAGCCGTG 120

QY 306 CCGTGGGCTCGGAAAGGAGATGAGAGGTGATCCGGGAGACAGCGTGCCTGGGCGGTTGCT 365
Db 121 CCGTGGGCTCGGAAAGGAGATGAGAGGTGATCCGGGAGACAGCGTGCCTGGGCGGTTGCT 180

QY 366 CGGAGACCCCGGACAGCTCGGTTGGCGGGACCATAGGCCGAAAGCACCCGGGAGCTCGTGG 425
Db 181 CGGAGACCCCGGACAGCTCGGTTGGCGGGACCATAGGCCGAAAGCACCCGGGAGCTCGTGG 240

QY 426 GCGAGCTGTGAGAGACAGCTCAGCGAAGAGAGAGAGCATTTGCGCTGGAAGTGG 485
Db 241 GCGAGCTGTGAGAGACAGCTCAGCGAAGAGAGAGAGCATTTGCGCTGGAAGTGG 300

QY 486 CCGTGGCTCTGGAAGAAGTGAAGGGCTGGGCGAGCGAGATGAGAAAGATGGAAGAG 545
Db 301 CCGTGGCTCTGGAAGAAGTGAAGGGCTGGGCGAGCGAGATGAGAAAGATGGAAGAG 360

QY 546 TGAAGCAACCTGCGCCAGAGCTCAAGAAAGAGTGAAGTTGAAGAGTCCAGTGGCAG 605
Db 361 TGAAGCAACCTGCGCCAGAGCTCAAGAAAGAGTGAAGTTGAAGAGTCCAGTGGCAG 420

QY 606 TGTTCATGAGAGAGAGACAGCCGCGAGAGTGAAGTTCACCTGTATGACTTTGACAA 665
Db 421 TGTTCATGAGAGAGAGACAGCCGCGAGAGTGAAGTTCACCTGTATGACTTTGACAA 480

QY 666 AGCGGCAAGGTCAACCGGAGAGGACATCAACAGCTTGTGACACATCATATAGAGTGG 725
Db 481 AGCGGCAAGGTCAACCGGAGAGGACATCAACAGCTTGTGACACATCATATAGAGTGG 540

QY 726 ACTCTCTGTCAACCACTCCCAACATCAAGCAAGATGCTGGGTTAAAGTCTCAACCGT 785
Db 541 ACTCTCTGTCAACCACTCCCAACATCAAGCAAGATGCTGGGTTAAAGTCTCAACCGT 600

QY 786 CCCCCGATGCGACGCAAGCAAGAGAGAGCTTGTTCATCAGGCTGACCTGCAAGCG 845
Db 601 CCCCCGATGCGACGCAAGCAAGAGAGAGCTTGTTCATCAGGCTGACCTGCAAGCG 660

QY 846 CAAAGCCCCGAGAGAGCAAGGCCCACTGAGGACCTGGGAGCTGGGAGAAAGAGAGC 905
Db 661 CAAAGCCCCGAGAGAGCAAGGCCCACTGAGGACCTGGGAGCTGGGAGAAAGAGAGC 720

QY 906 GAGGCCGCTCAGAGTTCCAGGGTGAAGCCGCTGAGCACTGTGCTGTAACCACT 965
Db 721 GAGGCCGCTCAGAGTTCCAGGGTGAAGCCGCTGAGCACTGTGCTGTAACCACT 780

QY 966 GCGTATGATGAGAAATGAGAGAGAAACCACTATGATCTGCGGGATAGAAAAC 1025
Db 781 GCGTATGATGAGAAATGAGAGAGAAACCACTATGATCTGCGGGATAGAAAAC 840

QY 1026 ACAGCTCCCAATTTGGGCTGGCTCCCTTCGTTGGGCGCAGAACTGAGCAAGATGCCCC 1085
Db 841 ACAGCTCCCAATTTGGGCTGGCTCCCTTCGTTGGGCGCAGAACTGAGCAAGATGCCCC 900

QY 1086 GCACTCTCAATCCCACTGATCTGCTCCATGAGCCGGAAGCATCCATCCACACC 1145
Db 901 GCACTCTCAATCCCACTGATCTGCTCCATGAGCCGGAAGCATCCATCCACACC 960

QY 1146 GAAAGCCCCAAGGGGTGAGACCCGGCTCTCTTCACTTCTTGAACAACCCCAATGCGCAAG 1205
Db 961 GAAAGCCCCAAGGGGTGAGACCCGGCTCTCTTCACTTCTTGAACAACCCCAATGCGCAAG 1020

QY 1206 TCTCAGAGCTCCAGCAACGGCTCCGGGGACCCAGAGCGGAGAGCACTTTGTGAGGT 1265
Db 1021 TCTCAGAGCTCCAGCAACGGCTCCGGGGACCCAGAGCGGAGAGCACTTTGTGAGGT 1080

QY 1266 CCCCCAAGGCCAGGCGAAGAGTGTGGTGGGCGCATGTTGCGCAAGAGGGCAGAAACA 1325
Db 1081 CCCCCAAGGCCAGGCGAAGAGTGTGGTGGGCGCATGTTGCGCAAGAGGGCAGAAACA 1140

QY 1326 AGCCCCCTCTGGAGACCCGCCATCCGTCGGGTGTCCTCCGCGCCACTGGCTGGCAGCC 1385
Db 1141 AGCCCCCTCTGGAGACCCGCCATCCGTCGGGTGTCCTCCGCGCCACTGGCTGGCAGCC 1200

QY 1386 CGGCGCTCTCTCCCTCTCTAGCCCTCCGAGCAAGAAAGCAACGAGCCAAAG 1445
Db 1201 CGGCGCTCTCTCCCTCTCTAGCCCTCCGAGCAAGAAAGCAACGAGCCAAAG 1260

QY 1446 AGAGCCAGAGGAGCTGCGGGGCTCTGAGGACCACTGGCTTGAAGTGGCTCTCTGG 1505
Db 1261 AGAGCCAGAGGAGCTGCGGGGCTCTGAGGACCACTGGCTTGAAGTGGCTCTCTGG 1320

QY 1506 GCGGGAGGACCTGCGGGAGCTGCGGCTTGTGTTGATGAGAGCCAGGCGGGGAGC 1565
Db 1321 GCGGGAGGACCTGCGGGAGCTGCGGCTTGTGTTGATGAGAGCCAGGCGGGGAGC 1380

QY 1566 CGGTCCAGAGACATGAGACCAACCAACCATGAACTCAACCACTTCACTTCT 1625
Db 1381 CGGTCCAGAGACATGAGACCAACCAACCATGAACTCAACCACTTCACTTCT 1440

QY 1626 ACCAGACA 1633
Db 1441 ACCAGACA 1448

RESULT 4
US-10-087-192-176
Sequence 176, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 176
LENGTH: 2807
TYPE: DNA
ORGANISM: Mus musculus
US-10-087-192-176

Query Match 60.8%; Score 1130.8; DB 13; Length 2807;
Best Local Similarity 84.6%; Pred. No. 4; 9e-260;
Matches 1311; Conservative 0; Mismatches 222; Indels 17; Gaps 3;

QY 282 GTGACAGCTTGGCGGTGAGGCTGCTGGGCTCGAAGAGGACATGAGAGTGGAG 341
Db 1 GTGACAGCTTGGCGGTGAGGCTGCTGGGCTCGAAGAGGACATGAGAGTGGAG 60

QY 342 GAGAGCGCTGCCGGGCGGTGTCTCGGACCCCGACAGCTGCGGTTGCGGGCACA 401
Db 61 GAGAGCGCTGTCCAGGAGGAGGTCTCAGAGACCCCGTGAAGTGGCAGGCACTG 120

QY 402 GCGGAGACCCCGGAGCTGTGGGCGAGCTGTGAGAGACAGCTCAGCGAGAGAG 461
Db 121 GCGGAGACCTCGGAGACCTGTGGGCTGACACTTCTAGAGAGGCTTGTGAGAGAG 180

QY 462 AGAGCACTTGGGCTGGAAGTGGCCCTGCTCTGTGAGAACTGACGGGCTGGGCAAGC 521
DB 181 AGGACGACTTCCCCCTGAGAGTGGCCCTGCGCTGAGAAATGACAGCCTAGTATG 240
QY 522 GAGATGAGAAAGATGAGAGAGTGAAGCAACCTGCGCCAGGGCTCCAGAAAGAGCTGA 581
DB 241 GAGATGAGAAAGATGAGAGAGTGAAGCAACCTGCGCCAGGGCTCCAGAAAGAGCTGA 300
QY 582 AGTTGAAAGAGCTTCAGTGGCAGCTGTCTCATGAGAGAGAGAGCCGGCAGAGTGAAGCT 641
DB 301 AGTTGAAAGAGCTTCAGTGGTGTCTCTGTGAGAGAGAGCAGCCGGCAGAGTGAAGCT 360
QY 642 TCACCTGTATGATTTGACAAACAAGAGTCAACCCGAGAGAGATCAACAGCTTGC 701
DB 361 TCACCTGTATGATTTGACAAACAAGAGTCAACCCGAGAGAGATCAACAGCTTGC 420
QY 702 TGACACCATATGATGAGTGTGAGCTCTCTGTCAACCACTCCCAACATCCAGACAGA 761
DB 421 TGACATCATATGATGAGTGTGAGCTCTCTGTGAGCACTTCCCAACATCCAGACAGA 480
QY 762 TGCTGCGGTTAAAGCTGACCGTGGCCCGATGGCAGCCAGACAGAGAGAGCGTCTTG 821
DB 481 CACTGCGGTTAAAGCTGACCGTGGCTCTGACGGGAGCCAGATTAAGAGAGCGTCTTG 540
QY 822 TCATTCAGGCTGACCTGCAAGCGGCAAGGCCCGAGACAGACCAAGCCCACTGAGAAC 881
DB 541 TCACACCTACGATCTGACAGACCAAGGCCCGAGACAGACCAAGCCCACTGAGAAC 600
QY 882 TGCGGAGCTGGGAGAGAGAGAGAGAGCGCCGCTCAGGTTCCAGGTTGACAGCGGCTTG 941
DB 601 TGCGGAGCTGGGAGAGAGAGAGAGAGCGCCGCTCAGGTTCCAGGTTGACAGCGGCTTG 660
QY 942 AGCAGTGTGCTGTACCACTTGCATGATGAGAACTGAGAGAGAGAGAAACCACTACT 1001
DB 661 AGCAGGCAAGCTGTATCACCACTTGCATGATGAGAACTGAGAGAGAGAGAAACCACTACT 720
QY 1002 TAGATCTGCGGAGATGAGAAAGCTACAGCTCCCAATTTGGGCTGGCTCCCTTCCGTG 1061
DB 721 TAGACCTGGGAGATGAGAAAGCTACAGCTCCCAATTTGGGAGTCCCGGATCCCTTCCGTG 780
QY 1062 CCCAGAGTCAAGACCTGCCCCCGGCACTTCATTCACACTGATTCGCTCCCATGAGC 1121
DB 781 CCCAGAGTCAAGACCTGCCCCCGGCACTTCATTCACACTGATTCGCTCCCATGAGC 840
QY 1122 CGGAGGCACTCACTCACTCCACACCGAAAGCCCAAGGCTGGAAGCCGCTCTCTTCACT 1181
DB 841 CAGAGGCTGCGCATCTCCACACCGGAGGCCCAAGGCTGGAAGCCGCTCTCTTCACT 900
QY 1182 TCCTTGAACCCCAATGCGCAAGGCTTCAAGCTTCAGCAACGGCTCCGGGGCAACCAAG 1241
DB 901 TCCTTGAACCCCAATGCGCAAGGCTTCAAGCTTCAGCAACGGCTCCGGGGCAACCAAG 960
QY 1242 ACCGGAGCAAGCACTTTGTGAGTCCCCCAAGGGCCAGGGGCAAGAGTGTGGGTC 1301
DB 961 ATGGGAGCAAGCACTTTGTGAGTCCCCCAAGGGCCAGGGGCAAGAGTGTGGGTC 1020
QY 1302 ACCTGCGCAGAGGGGCAAGAAACAAGCCCTCTTGGGACCCGCACTCTGCGGTGTCC 1361
DB 1021 ACCGGGCGCAGAGGGGCAAGAAACAAGCCCTCTTGGGACCCGCACTCTGCGGTGTCC 1080
QY 1362 CTTTCGCGCACTTGGCTGCGCAGCCGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1421
DB 1081 CTTTCGCGCACTTGGCTGCGCAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
QY 1422 AGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1481
DB 1141 AGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1200
QY 1482 TGGGCTCAGGTG--CCTGTCTCTGAGGCGGAGCACTTGGGAGGTGCTGCTTGG 1538
DB 1201 TGGGCTCAGGTGCTTCAAGCTGATGGGGCGGAGCAAGGTGAGGAGTGTGCTGCTG 1260
QY 1539 TGGGTATGAGAGCGAGCGCGGCGCTGAGAGACATGAGACCAACCAACCATG 1598

DB 1261 TGGGTATGAGAGCGAGCTGGGCAAGCCGCTGCAAGACAGAAACCATCACCACAGC 1320
QY 1599 AACATCACCAACCATTTACCACTTCTACAGACATAGAGCCCTCCCAAGGGCCCAAC 1658
DB 1321 AACATCACCAACCATTTACCACTTCTACAGACATAGAGCCCTCCCAAGGGCCCAAC 1369
QY 1659 CTGCAATGAG 1718
DB 1370 CTGCAAGGAG 1426
QY 1719 TTATTATGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1778
DB 1427 TTATTATGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1486
QY 1779 ACATGATGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1828
DB 1487 TCATGATGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1536

RESULT 5
US-09-993-966-2
Sequence 2, Application US/09993966
Publication No. US20030186232A1
GENERAL INFORMATION:
APPLICANT: ROHAN, MICHAEL
TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
FILE REFERENCE: 014024/028073
CURRENT APPLICATION NUMBER: US/09/993,966
CURRENT FILING DATE: 2001-11-27
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/252,884
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/325,571
PRIOR FILING DATE: 2001-10-01
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1416
TYPE: DNA
ORGANISM: Mus sp.
US-09-993-966-2
Query Match 57.9%; Score 1076.8; DB 10; Length 1416;
Best Local Similarity 85.5%; Pred. No. 3,1e-247;
Matches 1211; Conservative 0; Mismatches 202; Indels 3; Gaps 1;
QY 224 ATGGGAAACTTCACTCAAGCCGCGCGCTGTGCAAGCGCAGAGAGAGAGAGAGAGAGT 283
DB 1 ATGGGAAACTTCACTCAAGCCGCGCGCTGTGCAAGCGCAGAGAGAGAGAGAGAGAGT 60
QY 284 GACAGCTTGGCGGAG 343
DB 61 GACAGCTTGGCGGAG 120
QY 344 CAGGCTGCGCGGAG 403
DB 121 CAGGCTGCTCAAG 180
QY 404 CGAAGCAAG 463
DB 181 CGAAGCAAG 240
QY 464 GAGGACTTGGCTGAG 523
DB 241 GAGGACTTGGCTGAG 300
QY 524 GATGAGAAAGATGAG 583
DB 301 GATGAGAAAGATGAG 360

Db 721 CAGCAGACCTGCTACCACTATGCTGTGATGAGAAACATGAGAGAAACCTACTACTA 780
Qy 1004 GATTCGCGCGGATAGAAAATCAACAGTCCCAATTGGGCTGTGCTCCCTTCCGTGGCC 1063
Db 781 GACCTGCGGGGATAGAACTACACGCTCAGTTTGGACCGGATCCCTTCCGTGGCC 840
Qy 1064 CAGAGTCAAGATCTGCCCCCCCCCACTCCCACTCCCACTGATCTCGTCCCTCCATGAGCG 1123
Db 841 CAGAGTCAAGATCTGCCCCCCCCCACTCCCACTCCCACTGATCTCGTCCCTCCATGAGCG 900
Qy 1124 GAAGCCATCCATCCCAACCGAAAGCCCAAGCGGTGACCGGCTCTCTTCACTTC 1183
Db 901 GAAGCTCCCACTCCCAACCGAAAGCCCAAGCGGTGACCGGCTCTCTTCACTTC 960
Qy 1184 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGCTCCGAGGACCCAGAC 1243
Db 961 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGCTCCGAGGACCCAGAT 1020
Qy 1244 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGGGCA 1303
Db 1021 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGGGCA 1080
Qy 1304 GTGAGCAGAGGGGCAAGAAACAGACCCCTCTGGGACCCGCTGAGCTCCCTGAGTCCGCC 1363
Db 1081 GGGAGCAGAGGGTCAAGAGAGAGAGCTCTGAGTCCCACTCCCACTACTCTCTCCGCC 1140
Qy 1364 TCCGCCCACTGCTGCGAGCCCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1423
Db 1141 TCTGCCATCTGCGCACAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
Qy 1424 AAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1483
Db 1201 AAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1260
Qy 1484 GCCTCAGAGTGG---CCCTGTCTGTGGGGGGAGGACCTGCGGAGTGGCCGCTTGGTG 1540
Db 1261 GCTGAGAGAGCTCACCCTGTCATGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
Qy 1541 GTGTATGAGAGCCAGGCGGGGAGCCGGTCCAGAGACATGAGACCAACACCATGAA 1600
Db 1321 GTGTATGAGAGCCAGGCGGGGAGCCGGTCCAGAGACCAACACCATGAA 1380
Qy 1601 CATCACCACCATTTACACAC 1621
Db 1381 CATCACCACCATTTACACAC 1401

RESULT 7
US-09-993-966-6
; Sequence 6, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
; FILE REFERENCE: 014024/028073
; CURRENT APPLICATION NUMBER: US/09/993,966
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-6

Query Match 57.4%; Score 1066.6; DB 10; Length 1401;

Best Local Similarity 85.6%; Pred. No. 8.4e-245;
Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;
Qy 224 ATGGGGAATCTTCACTCCAAAGCCGCGCGTGTGCAAGGCAAGGAGAGCCGGAAGT 283
Db 1 ATGGGGAATCTTCACTCCAAAGCCGCGCGTGTGCAAGGCAAGGAGAGCCGGAAGT 60
Qy 284 GACAGCTTCCCGTGAAGCTGCTGCTGGGCTGGAGGCAATCGAGAGTGGATGGGAGA 343
Db 61 GACAGCTTCCCGTGAAGCTGCTGCTGGGCTGGAGGCAATCGAGAGTGGATGGGAGA 120
Qy 344 CAGCGTCCCGCGGCGGTCTGCGGAGCCCGACAGCTGGGCTGGGCGGACCATAGGC 403
Db 121 CAGCGTCCCGCGGCGGTCTGCGGAGCCCGACAGCTGGGCTGGGCGGACCATAGGC 180
Qy 404 CGAAGCAGCCGAGAGCTGCTGGGCGAGCTGTGAGAGACAGCTCAGAGAGAGAGAG 463
Db 181 CGAAGCAGCTGGGAGAGCTGCTGGGCGAGCTGTGAGAGAGAGCTCAGAGAGAGAG 240
Qy 464 GACGACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAGACTGACGGGCTGGGCGGAG 523
Db 241 GACGACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAGACTGACGGGCTGGGCGGAG 300
Qy 524 GATGAGAAAGATGAGAGAGTGAAGAAACCTGCGCCAGGCTCCAGAAAGCAGCTGAG 583
Db 301 GATGAGAAAGATGAGAGAGTGAAGAAACCTGCGCCAGGCTCCAGAAAGCAGCTGAG 360
Qy 584 TTGGAAGCTCAGTGCAGCTGTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 643
Db 361 TTGGAAGCTCAGTGTGATGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy 644 ACCCTGATGACTTTGACAAACAGGCAAGTCAACCGAGAGAGATCAACAGCTTGTG 703
Db 421 ACTTATATGACTTTCAGAAACATGCAAGTGCACCGTGAAGACATTAACAGCTTGTG 480
Qy 704 CACACCATATGAGAGTGTGAGCTCTCTGTCAACCACTCCCAATCAGCAAGATG 763
Db 481 CATACCATATGAGAGTGTGAGCTCTCTGTGACATTCCTCCCAATCAGCAAGATG 540
Qy 764 CTGCGGTAAGCTCAACCTGTGCGCCCGATGAGAGAGAGAGAGAGAGAGAGAGAG 823
Db 541 CTGCGGTAAGCTCAACCTGTGCGCCCGATGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 824 AATCAGGCTGACCTGCAAGCGCAAGGCCCTCCAGAGAGAGAGAGAGAGAGAGAG 883
Db 601 AATCAGGCTGACCTGCAAGCGCAAGGCCCTCCAGAGAGAGAGAGAGAGAGAGAG 660
Qy 884 CGAGCTGGAG 943
Db 661 CGAGCTGGAG 720
Qy 944 CAGCTGCTGCTTACCACTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1003
Db 721 CAGCTGCTGCTTACCACTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 1004 GATTCGCGCGGATAGAAAATCAACAGTCCCAATTGGGCTGTGCTCCCTTCCGTGGCC 1063
Db 781 GACCTGCGGGGATAGAACTACACGCTCAGTTTGGACCGGATCCCTTCCGTGGCC 840
Qy 1064 CAGAGTCAAGATCTGCCCCCCCCCACTCCCACTCCCACTGATCTCGTCCCTCCATGAGCG 1123
Db 841 CAGAGTCAAGATCTGCCCCCCCCCACTCCCACTCCCACTGATCTCGTCCCTCCATGAGCG 900
Qy 1124 GAAGCCATCCATCCCAACCGAAAGCCCAAGCGGTGACCGGCTCTCTTCACTTC 1183
Db 901 GAAGCTCCCACTCCCAACCGAAAGCCCAAGCGGTGACCGGCTCTCTTCACTTC 960
Qy 1184 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGCTCCGAGGACCCAGAC 1243
Db 961 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGCTCCGAGGACCCAGAT 1020
Qy 1244 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGAGTGGGTGGGCGAC 1303

Db 1021 GGAGAGAGAGCTTGTGAGGTGTCCTCCCAAGGCCAAGGCAAGATGGATGGGCGAC 1080
QY 1304 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGAGCCCGCATCTCGGTGTCCCCC 1363
Db 1081 GGGGCCAGAGGTGCAAGAGAGAGCTCACTGTACCCACCACTACTGTCTCCCC 1140
QY 1364 TTGGCCACCTGTGTCAGCGCCCTCTCTCCCTCCAGGCCCCCTGCGGACAAAG 1423
Db 1141 TTGGCCATCTGGCCACAGCCCTCTCTCCCATCTGGCACTCCCTGGGACAAAG 1200
QY 1424 AAGCACAAGACCCAGCCCAAGAGAGCCAGCAGGCTGCGGGGCTGCAAGCACTG 1483
Db 1201 AAACACAAGATGAGCCCAAGAGAGCCAGCTGCGGGGCTGCAAGGGGCCCTG 1260
QY 1484 GCCTCAGGTG--CCCTGTCTGGGGGGAGAGCACTGCGGGAGCTGCGGCTTGGTG 1540
Db 1261 GCTGCAAGAGGCTTCACCGTCTATGAGGGGAGAGGAGGAGGCTGCGGCTG 1320
QY 1541 GTGTATGAGAGCCAGCGGGGAGCGGATCCAGAGATGAGCACAACACCATGAA 1600
Db 1321 GTGTACAGAGCCAGGCTGGGAGGCGGTCCAGAGACAGAACCATCACCAACCA 1380
QY 1601 CATCACCACCATTTACACAC 1621
Db 1381 CATCACCACCATTTACACAC 1401

RESULT 8

US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Query Match 42.1%; Score 783; DB 10; Length 2379;

Best Local Similarity 100.0%; Pred. No. 7.9e-177;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 GGGTCCCTTCCTGCGGCCCAAGATCAGAACTGCCCCCGGCACCTCCATCCACTCGA 1105
Db 840 GGGTCCCTTCCTGCGGCCCAAGATCAGAACTGCCCCCGGCACCTCCATCCACTCGA 899
QY 1106 TCTGCTCCCATGAGCGGAGAGCCATCCATCCCAACCGAAAGCCCAAGGGGTGAC 1165
Db 900 TCTGCTCCCATGAGCGGAGAGCCATCCATCCCAACCGAAAGCCCAAGGGGTGAC 959
QY 1166 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACG 1225
Db 960 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACG 1019
QY 1226 CTCCGGGGCAACCCAGAGCGGAGCAAGCACTTTGTAGGTCTCCCAAGGCCCAAGGCAAG 1285
Db 1020 CTCCGGGGCAACCCAGAGCGGAGCAAGCACTTTGTAGGTCTCCCAAGGCCCAAGGCAAG 1079
QY 1286 AGTGTGAGTGTGGGCAAGTGGGCAAGGAGGAGCAAGAAACAAGCCCTCTGAGACCGGC 1345
Db 1080 AGTGTGAGTGTGGGCAAGTGGGCAAGGAGGAGCAAGAAACAAGCCCTCTGAGACCGGC 1139
QY 1346 ATCCCTGCGGTGTCCTCCCTCCGCCCACTTGCTGCGCAAGCCCGAGGCTCTCTCCCTA 1405
Db 1140 ATCCCTGCGGTGTCCTCCCTCCGCCCACTTGCTGCGCAAGCCCGAGGCTCTCTCCCTA 1199

QY 1406 GCCCCCTCGGGGCAAGAGCAAGAGCAACCGACCCAGAGGAGGAGGAGGAGGCTGCGG 1465
Db 1200 GCCCCCTCGGGGCAAGAGGAGCAAGAGCAACCGACCGAGCCAGAGGAGGAGGAGGCTGCGG 1259
QY 1466 GGCCTGAGGAGCAACACTGAGCTGAGGTGCGCTGTCTGTGGGGGAGAGCACTGCGGAG 1525
Db 1260 GGCCTGAGGAGCAACACTGAGGTGCGCTGTCTGTGGGGGAGAGCACTGCGGAG 1319
QY 1526 CTGCGGCTTGGTGTGTATGAGAGCGAGCGGGGAGCGGATCCAGAGCATGAGAC 1585
Db 1320 CTGCGGCTTGGTGTGTATGAGAGCGAGCGGGGAGCGGATCCAGAGCATGAGAC 1379
QY 1586 CACCAACCATTAATATGACCAACCATTAACCACTTCTTACAGAGCATGAGAGCCCTCC 1645
Db 1380 CACCAACCATTAATATGACCAACCATTAACCACTTCTTACAGAGCATGAGAGCCCTCC 1439
QY 1646 CAGAGGCCCACTGTCATATGAGAGCCCAACCCCGACACCAACAGCATTTATTTAT 1705
Db 1440 CAGAGGCCCACTGTCATATGAGAGCCCAACCCCGACACCAACAGCATTTATTTAT 1499
QY 1706 CTATTAATTAATGTTATGATGATGATTTATTTATTAATTTATTTACTCACTAAT 1765
Db 1500 CTATTAATTAATGTTATGATGATGATTTATTTATTAATTTATTTACTCACTAAT 1559
QY 1766 ATTTAGCTAGCTTACATGTAAGATCTATGAAACAGAGCACTAACTTTATTTAT 1825
Db 1560 ATTTAGCTAGCTTACATGTAAGATCTATGAAACAGAGCACTAACTTTATTTAT 1619
QY 1826 GTT 1828
Db 1620 GTT 1622

RESULT 9

US-09-764-891-5904
; Sequence 5904, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 5904
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5904

Query Match 42.1%; Score 783; DB 10; Length 2379;

Best Local Similarity 100.0%; Pred. No. 7.9e-177;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 GGGTCCCTTCCTGCGGCCCAAGATCAGAACTGCCCCCGGCACCTCCATCCACTCGA 1105
Db 840 GGGTCCCTTCCTGCGGCCCAAGATCAGAACTGCCCCCGGCACCTCCATCCACTCGA 899
QY 1106 TCTGCTCCCATGAGCGGAGAGCCATCCATCCCAACCGAAAGCCCAAGGGGTGAC 1165
Db 900 TCTGCTCCCATGAGCGGAGAGCCATCCATCCCAACCGAAAGCCCAAGGGGTGAC 959
QY 1166 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACG 1225
Db 960 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACG 1019
QY 1226 CTCCGGGGCAACCCAGAGCGGAGCAAGCACTTTGTAGGTCTCCCAAGGCCCAAGGCAAG 1285
Db 1020 CTCCGGGGCAACCCAGAGCGGAGCAAGCACTTTGTAGGTCTCCCAAGGCCCAAGGCAAG 1079

Qy	1286	AGTGTGGGTGTGGGCGACAGTGGCCAGAGGGCGAAGAAACAAGCCCTCTGGGACCCGCG	1345
Db	1080	AGTGTGGGTGTGGGCGACAGTGGCCAGAGGGCGAAGAAACAAGCCCTCTGGGACCCGCG	1139
Qy	1346	ATCCCTGGGTGTGTCCTCCCTCTCGGCCACCTGGCTGCGAGCCCGGCGCTCTCCCTCCCTTA	1405
Db	1140	ATCCCTGGGTGTGTCCTCCCTCTCGGCCACCTGGCTGCGAGCCCGGCGCTCTCCCTCCCTTA	1199
Qy	1406	GCCCCCTCTGGGGCACAAGACACAAGCACTGAGCGAAGAGAGCGCCACACAGGGCTGCGGG	1465
Db	1200	GCCCCCTCTGGGGCACAAGACACAAGCACTGAGCGAAGAGAGCGCCACAGGGCTGCGGG	1255
Qy	1466	GGCCTGCGAGCACACTGGCTCAAGTGGCCCTGTCTTGGGCGGGAGCACACTGGCGGAG	1522
Db	1260	GGCCTGCGAGCACACTGGCTCAAGTGGCCCTGTCTTGGGCGGGAGCACACTGGCGGAG	1319
Qy	1526	CTGCCCGCTTGTGTGTGTATGAGAGCCAGGCCGGGCGACCGGCTCAGAGACATGAGCAC	1585
Db	1320	CTGCCCGCTTGTGTGTGTATGAGAGCCAGGCCGGGCGACCGGCTCAGAGACATGAGCAC	1379
Qy	1586	CACCAACCACTGAACATCACCACTTACCACTTCTTACAGACATAGAGCCCTTCC	1645
Db	1380	CACCAACCACTGAACATCACCACTTACCACTTCTTACAGACATAGAGCCCTTCC	1439
Qy	1646	CCAGGGCCCCCAGCCGCAATGAAAGAGCCCAACCCCGGACACCAAGGCACTTATTTT	1705
Db	1440	CCAGGGCCCCCAGCCGCAATGAAAGAGCCCAACCCCGGACACCAAGGCACTTATTTT	1499
Qy	1706	CTATTAAATTATGTATTATGATGATTAATTGTTATTAATAATTATTTACTTCGACTAAT	1765
Db	1500	CTATTAAATTATGTATTATGATGATTAATTGTTATTAATAATTATTTACTTCGACTAAT	1555
Qy	1766	ATTTAAGCTAGCCTACATGTAGAAAGTCTATGGAACACAGAACTTAACTTTTATTAAT	1822
Db	1560	ATTTAAGCTAGCCTACATGTAGAAAGTCTATGGAACACAGAACTTAACTTTTATTAAT	1619
Qy	1826	GTT 1828	
Db	1620	GTT 1622	

```

RESULT 10
US-10-087-192-178
Sequence 178, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David M.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/7747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 178
LENGTH: 106315
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) _ (106315)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-178

```

Query Match	42.1%	Score 783	DB 13	Length 106315
Best Local Similarity	100.0%	Pred. No. 4.1e-176		
Matches 783; Conservative	0	Mismatches 0	Indels 0	Gaps 0

OY	1046	GGCTCCCTTCGAGCCAGAAAGTCAGAACTGGCCCCCGCAGCTCCTCAATCCCACTCGA	1105
Db	95728	GGCTCCCTTCGAGCCAGAAAGTCAGAACTGGCCCCCGCAGCTCCTCAATCCCACTCGA	9578
OY	1106	TTCTGCTTCCATGAGCCGGAAAGCCATCCACATCCACACCGAAAGCCCAAGCGCTGAGC	1165
Db	95788	TTCTGCTTCCATGAGCCGGAAAGCCATCCACATCCACACCGAAAGCCCAAGCGCTGAGC	9584
OY	1166	CCGGCCCTCTTCCACTTCTCTGACACCCCAATGSCCAAGGCTCAGAGCTCCAGCAACGG	1225
Db	95848	CCGGCCCTCTTCCACTTCTCTGACACCCCAATGSCCAAGGCTCAGAGCTCCAGCAACGG	9590
OY	1226	CTCCGGGGGCAACCCAGAGCGGGAGCAACACTTTGTGAGTCTCCCAAGGCCCAAGGCGAG	1285
Db	95908	CTCCGGGGGCAACCCAGAGCGGGAGCAACACTTTGTGAGTCTCCCAAGGCCCAAGGCGAG	9596
OY	1286	AGTGTGGGTGTGGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGCC	1345
Db	95968	AGTGTGGGTGTGGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGACCCGCC	9602
OY	1346	ATCCCTGAGGAGTATCCCCCTCGGCCCACTGAGCTGCAGCCCGGCCCTCTCCCTCCCTA	1405
Db	96028	ATCCCTGAGGAGTATCCCCCTCGGCCCACTGAGCTGCAGCCCGGCCCTCTCCCTCCCTA	9608
OY	1406	GCCCCCTCGGGCAAGAAAGCAACAGCACCGAGCCAGAGAGAGCCAGAGGCTGCGG	1465
Db	96088	GCCCCCTCGGGCAAGAAAGCAACAGCACCGAGCCAGAGAGAGCCAGAGGCTGCGG	9614
OY	1466	GGCTGTGAGGCAACCACTGGCTCTCAGGTGGGCTCTCTCTTGGGGGGGAGCACTCGGGGAG	1525
Db	96148	GGCTGTGAGGCAACCACTGGCTCTCAGGTGGGCTCTCTCTTGGGGGGGAGCACTCGGGGAG	9620
OY	1526	CTGCCCGCCTTGTGTGTGTATGAGAGCCAGAGCGGGCAGCGGTCCAGAGCACTGAGCAC	1585
Db	96208	CTGCCCGCCTTGTGTGTGTATGAGAGCGGGCAGCGGTCCAGAGCACTGAGCAC	9626
OY	1586	CACCAACAACATGAAACATCACACCATTTACCACTTTCTACAGACATAGAGCCCTCC	1645
Db	96268	CACCAACAACATGAAACATCACACCATTTACCACTTTCTACAGACATAGAGCCCTCC	9632
OY	1646	CCAGGGGCCCAACCCCTGCATATGAAAGACCCCAACCCCGGACACCAAGGATTAATTAAT	1705
Db	96328	CCAGGGGCCCAACCCCTGCATATGAAAGACCCCAACCCCGGACACCAAGGATTAATTAAT	9638
OY	1706	CTATTAAATTAATGTTATTAATGATGATTAATTTGTTAATTAATTAATTTGTTACTCCATTAAT	1765
Db	96388	CTATTAAATTAATGTTATTAATGATGATTAATTTGTTAATTAATTAATTTGTTACTCCATTAAT	9644
OY	1766	ATTTAGCTAGCTTACATGTGAGAGATCTATGAAACACAGACTTAACTTTTATTTAT	1825
Db	96448	ATTTAGCTAGCTTACATGTGAGAGATCTATGAAACACAGACTTAACTTTTATTTAT	9650
OY	1826	GTT 1828	
Db	96508	GTT 96510	

```

RESULT 11
US-09-993-966-21
; Sequence 21, Application US/09993966
; Publication No. US20030186332A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/328073
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571

```



```
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8686
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(427)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8686

Query Match      22.5%; Score 418.2; DB 10; Length 427;
Best Local Similarity 99.3%; Pred. No. 7,7e-90;
Matches 420; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1076 CTGCCCCCGGCACTTCATCCCATCTGATCTGCTCCCATGAGCCGGAAGCCATCCAC 1135
DB 4 CTGCCCCCGGCACTTCATCCCATCTGATCTGCTCCCATGAGCCGGAAGCCATCCAC 63

QY 1136 ATCCCAACGGAAGAGCCCAAGGGGTGAGACCCGGGCTCTTCCACTTCTTGAACCCCA 1195
DB 64 ATCCCAACGGAAGAGCCCAAGGGGTGAGACCCGGGCTCTTCCACTTCTTGAACCCCA 123

QY 1196 ATGCCAAGGTCTTCAAGCTTCCAGCAAGGCTCCGGGCAACCAAGAGAGCAAGCAC 1255
DB 124 ATGCCAAGGTCTTCAAGCTTCCAGCAAGGCTCCGGGCAACCAAGAGAGCAAGCAC 183

QY 1256 TTTGTGAGGTCTCCCAAGGCGCCAGAGGCAAGAGTGTGGGTGGGCGCACGTGGCCAGAGG 1315
DB 184 TTTGTGAGGTCTCCCAAGGCGCCAGAGGCAAGAGTGTGGGTGGGCGCACGTGGCCAGAGG 243

QY 1316 GCAAGAAACAGAGCCCTCTGGAGACCGGCATCCCTGGGATCCCTCCGCCACCTG 1375
DB 244 GCAAGAAACAGAGCCCTCTGGAGACCGGCATCCCTGGGATCCCTCCGCCACCTG 303

QY 1376 GCTGCCAGCCGCGCTCTCCCTCTAGCCCTCCGAGCAAGAAAGCAAGCAC 1435
DB 304 GCTGCCAGCCGCGCTCTCCCTCTAGCCCTCCGAGCAAGAAAGCAAGCAC 363

QY 1436 CGAGCCAGAGAGAGCCAGCGAGGTGCGGGGCTTGAAGCACTGGCTTCAAGTGC 1495
DB 364 CGAGCCAGAGAGAGCCAGCGAGGTGCGGGGCTTGAAGCACTGGCTTCAAGTGC 423

QY 1496 CCT 1498
DB 424 CCT 426

RESULT 14
US-09-764-868-152
; Sequence 152, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152
```

```
; LENGTH: 1966
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-152

Query Match      14.7%; Score 272.4; DB 9; Length 1966;
Best Local Similarity 56.5%; Pred. No. 1.2e-54;
Matches 667; Conservative 0; Mismatches 471; Indels 42; Gaps 7;

QY 497 GAGAAAGCTGACGGGCTGGGACGGAGATGAGAAAGATGAGAGAGTGAAGAACCC 556
DB 516 GAGAACCCGGGAGCAACTCTCAGCGAGATGACGAGAGAGGGGAGCAAAACCGGAGGCG 575

QY 557 TGCCCAAGCTC---CAAGAGCAGCTGAATTGGAAGACTCCAGTCCGACGTCCATG 613
DB 576 CCGGAGAGACCGGGCGGGCGCCCTTCATTTGACGCACTTCCAGTCCGATGCTCCGTG 635

QY 614 GAGAGAGACAGCCGGGACGAGTGAACCTTCACTTGAATGACTTTGACAAACGGCAG 673
DB 636 GAGAGAGACAGCCGGGACGAGTGAACCTTCACTTGAATGACTTTGACAACTGGGAG 695

QY 674 GTCAACCGAGAGACATCAACAGCTTGTGACACCATCTATGAGGTGTGACTCTCT 733
DB 696 GTCAACCGAGAGACATCAACAGCTTGTGACACCATCTATGAGGTGTGACTCTCTG 755

QY 734 GTCAACCATCTCCCAACATCAAGAGATCTGGGGTAAAGCTCACCGTGGCCCGGAT 793
DB 756 GTCAACCATCTCTGGGACAGACAGACCCCTCGGTGTGAAGCTTAACCGTCAAGCCCTG 815

QY 794 GGCAGCCAGAGCAAGAGAGCGTCTTGTCAATCAGGCTGACCTGACGAGAGGCAAGGCC 853
DB 816 CCTTCCAGCAAGAGAGAGAGGCTCTCTCTGTGCGAGAGACCGGAGAGCCACCGCTTGC 875

QY 854 CGAGCAGAGACCAAGCCCACTGAGAGACTTGGAGCTGGGAGAGAGAGAGAGAGCCCG 913
DB 876 AGATGAGAGGTGAATCTGAGAGAGAGAGCAAGGAGGTGCTGAGAGAGGTGTCTGCACAC 935

QY 914 CTGAGGTTCAGAGGTGACAGCCGCTGTGAGCACTGTGCTGCTACCACTTGTGCTGAT 973
DB 936 GTCAAGAGGCGCAGTACTGACCCCAAGCCCTGTGCGAGAGGAGGCGCTCTCTGCTGAG 995

QY 974 GAGAACTGAGAGAGAGAAACCACTTATGATCTGCGGGGATGAGAAACTACAGTCC 1033
DB 996 GAGAACTGAGAGAGAGAGAAACCACTTATGATCTGCGGGGATGAGAAACTACAGTCC 1055

QY 1034 CAATTGGGCTTGGCTTCCCTTCCGTGAGCCAGAGATCAAGACTGCCCCCGCACCTCC 1093
DB 1056 AGATTGGGCTTGGGCTTCCCTTCCGTGAGCAAGAGAGAGAGCCAGAGGAGGCTCCG 1115

QY 1094 AATCCCA---CTGATCTGCTCTCCATGAGCCGAAAGCATTCACATCCACACCGAAAG 1150
DB 1116 CACTTCCAGGCGCGGCTCCGCTCCAGAGCCAGATACATCCGATACACACCGGAGG 1175

QY 1151 CCCCAAGGCGTGAACCGGCTCTCTTCACTTGTGACACCGCCATTCGCAAGGTCTCA 1210
DB 1176 TCAAGGTGCTGTGTAACAAGTGTGCAAGCTTCCGA---GCTGTGCTCGGAGCTCTG 1232

QY 1211 GAGCTCAGCAACGAGCTCCGAGGACCCAGAGAGGAGCAAGCACTTGTGAGTCCCGCC 1270
DB 1233 GACACGAGCCCGGCGGAGAGGCGGAG-----AAGCAGTTCTCAAGTCCCGCC 1283

QY 1271 AAGGCCAGGAGCAAGAGTGTGTGAGGCTACGTGCGCAGAGGGGCAAGAAACAGCCC 1330
DB 1284 AAGGGCTCCGAGAGAGCGCTCGGAGGTCAGACGAGCAGCAAGTCCGAGAAAGCTTCAAG 1343

QY 1331 CTTCTGAGACCCGCAATCCCTGCGGTGCTCCCGCCCACTGTGCTGACGCGGCGC 1390
DB 1344 TACTTACTGCTCGGCGCTCC-----TCCGCTCCAGGCTCTTGAAGCGCACACCTC 1397

QY 1391 CTCTCCCTCTCTTACCCCTCCGAGCAAGAGAGCAAGAGCAAGAGCAAGAGAGAG 1450
DB 1398 CCGAGAGCCCAACCGCACCTTACGAGCACAAGAGGCTACCGCAAAAGGAGAGAGGCG 1457
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 8793 Seconds

(without alignments)
7704.022 Million cell updates/sec

Title: US-09-993-966-5

Sequence score: 1859
1 gaattcgcccttctaacy.....aaaaaaaaaagcgccgc 1859

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_esc1.*
2: gb_esc2.*
3: gb_hic.*
4: gb_esc3.*
5: gb_esc4.*
6: gb_esc5.*
7: gb_esc6.*
8: gb_gsa1.*
9: gb_gsa2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1753.2	94.3	2142	3	AF289584 Homo sapi
2	1242.8	66.9	2289	3	AK082367 Mus muscu
3	1197.2	64.4	2471	3	AK076164 Mus muscu
4	1027.6	55.3	1466	3	AK013241 Mus muscu
5	885	47.6	885	9	AY412098 Homo sapi
6	826.8	44.5	921	5	BQ645656 AGENCOURT
7	825.6	44.4	931	5	BQ653673 AGENCOURT
8	810.6	43.6	990	5	BQ064678 AGENCOURT
9	793	42.7	928	5	BQ064956 AGENCOURT
10	789.4	42.5	895	5	BQ644360 AGENCOURT
11	757.6	40.8	953	5	BQ644360 AGENCOURT
12	748.8	40.3	885	9	AY412099 Pan trogl
13	712.6	38.3	966	5	BQ649813 AGENCOURT
14	711.4	38.0	1008	5	BQ652087 AGENCOURT
15	706	38.0	1008	5	BQ646371 AGENCOURT
16	700.4	37.7	1027	4	BG104777 602311726
17	689.6	37.1	888	9	AY412100 Mus muscu
18	682.6	36.7	1059	5	BQ070932 AGENCOURT
19	661.6	35.6	825	6	CB961961 AGENCOURT
20	622	33.5	622	6	CB215756 NISC np09
21	611.8	32.9	899	5	BQ645507 AGENCOURT
22	597.8	32.2	826	6	CB961668 AGENCOURT
23	593.6	31.9	913	5	BQ895164 AGENCOURT
24	593	31.9	634	7	CN369119 170005999

25	592.2	31.9	777	4	BG542261 602571809
26	584.2	31.4	627	4	B1767278 603057995
27	583.2	31.4	915	6	CB182455 AGENCOURT
28	561.4	30.2	564	4	BM711145 UI-E-DX1-
29	549.2	29.5	654	9	AG112447 Pan trogl
30	532	28.6	677	7	CF731367 UI-M-G20-
31	530	28.5	674	6	CB056689 NISC j119
32	509.2	27.4	946	6	BY712252 BY712252
33	508.2	27.3	724	6	CA328380 UI-M-FY0-
34	503.8	27.1	839	7	CK795599 AGENCOURT
35	477.4	25.7	751	4	BG820139 602782356
36	467.4	25.1	481	4	BT047069 MR3-FN020
37	466.2	25.1	678	2	BF537335 602053266
38	459.4	24.7	653	2	BF540602 602054683
39	458	24.6	645	5	BQ554861 H4030603-
40	456.6	24.6	670	2	BB650914 BB650914
41	455.4	24.5	843	7	CO811260 AGENCOURT
42	452.6	24.3	693	5	BM935129 UI-M-BH3-
43	451.2	24.3	931	5	BQ856404 AGENCOURT
44	431	23.2	817	7	CK790579 AGENCOURT
45	425.8	22.9	429	7	CN369118 170005322

ALIGNMENTS

RESULT 1	AF289584	2142 bp	MRNA	linear	HTC 01-JAN-2002
LOCUS	AF289584				
DEFINITION	Homo sapiens clone pp7246 unknown mRNA.				
ACCESSION	AF289584				
VERSION	AF289584.1	GI:18027371			
KEYWORDS	HTC.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2142) Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T., Wan, D.F. and Gu, J.R.				
TITLE	Novel human cDNA clones with function of inhibiting cancer cell growth				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2142) Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T., Wan, D.F. and Gu, J.R.				
AUTHORS	Man, D.F. and Gu, J.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Jn 2200 Xie-Tu Road, Shanghai 200032, P. R. China				
FEATURES	Location/Qualifiers				
source	1. 2142				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="pp7246"				
	277. 1689				
CDS	/product="unknown"				
	/protein_id="AA155768.1"				
	/db_xref="GI:18027372"				
	/translation="MGKLSKPDVCKRRRESPEGDSFAVSAMARKGIEWIGRQCPGGVSGRQRLAGTIGRSTREIVGLDRLTLEBEDDRLLEVALPEPEKTVGDSGDEKKMERSEPCPSGSKQLEKEELQCDVSMEDSRQEWFTFLYFDNNKAGTRBEDITSLLHTIYEVDSVNHSPSTSSKMLVYKLVAVAGDSQSKSVVNOADLOSAPRATKTEBDRSEWKORAPLRQGSRLRQSGCYHNCDEBNTERRHYVDIAGIENVTSGFGSPSVAQSELPPTSNPTRSRSHPEPAIHHPKPGQVDPAFRHFDTPIAKUSELQORLGTQDSGRFVRSPPVAQGSVGVGVAVAGANKKPLGAIIPAVSSAHLAASPALPSLAPLHGKHKRAKESQCGCRGLQAPLASGSPVLRERHRELIPALVVESSAQGPVQRIHHHHHHHHHHFYQT"				
ORIGIN					
Query Match	94.3%	Score 1753.2	DB 3	Length 2142	

Best Local Similarity 99.5%; Pred. No. 0;
Matches 1750; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

QY 63 GGAAGTCCGAGCCGCGAGAGCGGCGGAGAGCGGTCGCCGCGCGCTTCGGGCTCCGCT 122
Db 116 GTGAGTCCGAGCCGCGAGAGCGGCGGAGAGCGGTCGCCGCGCGCTTCGGGCTCCGCT 175
QY 123 CGGCTCGGGGGCTTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 182
Db 176 CGGCTCGGGGGCTTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 235
QY 183 GCGCATGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
Db 236 GCGCATGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295
QY 243 AGCGGCGCGCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
Db 296 AGCGGCGCGCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355
QY 303 CTGCGTGGGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
Db 356 CTGCGTGGGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 415
QY 363 TCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
Db 416 TCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
QY 423 TGGGCGAGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
Db 476 TGGGCGAGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
QY 483 TGGGCGCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542
Db 536 TGGGCGCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 595
QY 543 GAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
Db 596 GAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655
QY 603 ACGTGTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
Db 656 ACGTGTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 715
QY 663 ACAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
Db 716 ACAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775
QY 723 TGAAGTCTCTGTCAACCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
Db 776 TGAAGTCTCTGTCAACCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 835
QY 783 TGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 842
Db 836 TGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895
QY 843 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
Db 896 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955
QY 903 AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 962
Db 956 AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1015
QY 963 ATTGCGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022
Db 1016 ATTGCGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075
QY 1023 ACTACAGTCCCAATTTGGGCGTGGCTCCCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082
Db 1076 ACTACAGTCCCAATTTGGGCGTGGCTCCCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1135
QY 1083 CCGGCACTCCCAATCCCACTGATCTGCGTCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1142

```

```

Db 1136 CCGGCACTCCCAATCCCACTGATCTGCTCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1195
QY 1143 ACCGAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1202
Db 1196 ACCGAAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1255
QY 1203 AGGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1262
Db 1256 AGGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1315
QY 1263 GATCCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1322
Db 1316 GATCCCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1375
QY 1323 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
Db 1376 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435
QY 1383 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442
Db 1436 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1495
QY 1443 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1502
Db 1496 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1555
QY 1503 TGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
Db 1556 TGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1615
QY 1563 AGCGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1622
Db 1616 AGCGGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
QY 1623 TCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1682
Db 1676 TCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
QY 1683 CGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742
Db 1736 CGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
QY 1743 ATATATATTTACTCCACTAATATTTAGCTAGCTACATGATGAGAGAGAGAGAGAGAGAGAG 1802
Db 1796 ATATATATTTACTCCACTAATATTTAGCTAGCTACATGATGAGAGAGAGAGAGAGAGAGAG 1855
QY 1803 CAGAGCTAAACTTTATTTATATGTT 1828
Db 1856 CAGAGCTAAACTTTATTTATATGTT 1881

```

```

RESULT 2
AK082367 2289 bp mRNA linear HTC 03-Apr-2004
LOCUS Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
DEFINITION enriched library, clone:C23004009 product:naked cuticle 1 homolog
(Drosophila), full insert sequence.
ACCESSION AK082367.1 GI:26100612
VERSION AK082367.1
KEYWORDS HTC; CAP triepiper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLES High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

```


TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBLISHED	20493374
REFERENCE	11042159
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaka, S., Sasaki, N., Carninci, P., Komano, H., Akiyama, J., Nishi, K., Katsunai, T., Teshiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasahagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBLISHED	20530913
REFERENCE	11076661
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2289)
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hizomoto, K., Hiroaka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komano, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sugabe, Y., Tagami, M., Tagawa, A., Takedashi, F., Takaku-Akai, S., Tanaka, Y., Tanaka, T., Tamaru, A., Toy, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-Apr-2002)
COMMENT	Yoshitake Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research@riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
FEATURES	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse libraries.
Source	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers 1. 2289 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM.DB:C230040009" /db_xref="taxon:T0090" /clone="C230040005" /tissue_type="cerebellum" /clone_type="RIKEN full-length enriched mouse cDNA library" /dev_stage="0 day neonate" 152_1568 /note="naked cucle 1 homolog (Drosophila)" [MG1:2135954, GB NM_027280, evidence: BLASTN, 98%, match:1658]

[illegible]

Db 953 CTAAGCTCAAGTTGGACCGGATCCCTTGCTGGTGCCCAAGATCAGAGCTGCCCC 1012
 Qy 1084 CCGCAGCTCCAAATCCCATCTGATTCGCTGCCATGAGCCGGAAGCCATCCATCCACA 1143
 Db 1013 TCGAATCTCAACCCCACTGCTCTCCCTCCACGAGCCAGAACTCCCAATCCACA 1072
 Qy 1144 CCGAAGACCCCAAGGCGTGGACCCGCGCTCTTCCATCTTCCTTGACACCCCAATCGCAA 1203
 Db 1073 CCGAGGCCCCCAAGGTGGACCCAGGCTCTTCCATCTCTTGACACCCCAATTTGCCAA 1132
 Qy 1204 GGTCTCAGAGCTCCAGCAAGCTCCGCGGACCCAGACCGGAGCAAGCACTTTGTAG 1263
 Db 1133 GGCATCAGAGCTCCAGCAAGCGCTCCGCGGACCTCAGATGGAGCAAGCACTTTGTAG 1192
 Qy 1264 GTCCCCCAAGGCCCCAAGAGTGTGGTGTGGGCAAGTGGCCAGAGGGGCAAGAA 1323
 Db 1193 GTCCCCCAAGGCCCCAAGAGTGTGGTGTGGGCAAGTGGGCAAGAGTGGCAAG 1252
 Qy 1324 CAAGCCCCCTCTGGGACCCGCGCATCCCTGGGCTGCCCTCCGCGCAAGCTGGCGAG 1383
 Db 1253 CAAGCTCCACTGTATCCCAACCACTCACTGTCTCCCTCTGCGCAATCTGGCCACAG 1312
 Qy 1384 CCGGCGCTCTCTCCCTCCCTAGCCCTCGGACCAAGAGCAAGCAAGCAAGCCAGCCAA 1443
 Db 1313 CCGAGCCCTTCTCCCACTGCGACCCCTGGGCAAGAAACAGAAAGATGAGCCAA 1372
 Qy 1444 GAGAGCCAGAGGCTGCGCGGCTGCGAGCACACTGGCCCTCAGAGTG--CCCTGT 1500
 Db 1373 GAGAGCCAGAGGCTGCGCGGCTGCGAGGCCCCCTGCGCAAGAGCTTCACCGT 1432
 Qy 1501 CCGGCGCGGAGACCTGCGGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1560
 Db 1433 CAGGCGCGGAGAGAGGTGAGAGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 1492
 Qy 1561 GCAAGCCGCTCAGAGACATAGACCAACCAACCATGAACATGACCAATTCACCA 1620
 Db 1493 GCAAGCCGCTCAGAGACATAGACCAACCAACCATGAACATGACCAATTCACCA 1552
 Qy 1621 CTCTTACCAACATAGAGCCCCCTCCAGGGCCCCACCTGCTGCTGCTGCTGCTGCTGCT 1680
 Db 1553 CTCTTACCAACATAGAGCCCCCTCCAGGGCCCCACCTGCTGCTGCTGCTGCTGCTGCT 1601
 Qy 1681 CCGGACCAAGAGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
 Db 1602 C---ACACCTTAAGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1658
 Qy 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
 Db 1659 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1718
 Qy 1801 CACGAGCTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1828
 Db 1719 CACGAGCTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1746

RESULT 3
 LOCUS AK076164 2471 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
 enriched library, clone:3230401D10 product:naked cuticle 1 homolog
 (Drosophila), full insert sequence.
 ACCESSION AK076164
 VERSION AK076164.1 GI:26096648
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Mammalia; Eutheria; Rodentia; Scuriongnathi; Muridae; Murinae; Mus.

REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253

PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komuro, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hazada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujimake, S., Inoue, K., Togawa, Y., Izawa, K., Tanaka, M., Ohara, E., Matsubara, S., Kawai, D.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, D.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 20530913
 MEDLINE 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 JOURNAL of 60,770 full-length cDNAs
 MEDLINE Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 2471)
 AUTHORS Aichi, U., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komuro, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Ohnato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/.
 URL: http://fantom.qualifiers

FEATURES
 source
 1..2471
 /organism="Mus musculus"
 /molecule="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM:DB:3230401D10"
 /db_xref="taxon:10090"
 /clone="3230401D10"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="14, 17 days embryo"

AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	
TITLE	
JOURNAL	Garninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
PUBMED	Normalizatization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)
AUTHORS	
TITLE	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20499374
REFERENCE	11042159
AUTHORS	
TITLE	
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Onura, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
PUBMED	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipicillary sequencer
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	
TITLE	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	20530913
REFERENCE	11076861
AUTHORS	
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL	Functional annotation of a full-length mouse cDNA collection
PUBMED	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
PUBMED	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 1466)
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Katao, H., Kawai, U., Koijima, Y., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
PUBMED	Direct Submission
REFERENCE	Submitted (10-JULY-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
AUTHORS	Please visit our web site (http://genome.gsc.riken.jp/) for further details.
TITLE	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.
JOURNAL	Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGACGATCCAGAGCTCTTTTCTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to Rot = 37.5. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGGAGATTTCGAGTTAATTAATATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SclI. Cloning sites, 5' end: XhoI, 3' end: SclI.

		Host: SOLR.	
FEATURES		Location/Qualifiers	
	source	1..1466 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:2810434J10" /db_xref="taxon:10090" /clone="2810434J10" /issue_type="whole body" /clove.lib="RIKEN full-length enriched mouse cDNA library" /dye_stage="10, 11 days embryo"	
misc_feature		1..1466 /note="naked cuticle 1 homolog (Proscophila) [MGD GI:2135954, GB NM_027280, evidence: BLASTN, 98%, match=1658]"	
ORIGIN			
Query Match	55.3%; Score 1027.6; DB 3; Length 1466;		
Best Local Similarity	84.7%; Pred. No. 7.Be-210;		
Matches 1194;	Conservative 0; Mismatches 199; Indels 17; Gaps 3;		
OY	415 GGAGCTGTTGGGGCCACCGTGTGAAGAGCACGCTCAGCGAGGAAAGAGAGCACTTCG	474	
Db	71 GGAATCTGGTGTAACACTTCTAGAAGGCTCTCGTAGAGAGACGAGCACTTCCC	130	
OY	475 GCTGGAATGGCCCTTCCTCTGGAAGACTGACGGCTGGGCAGCGGAGTAGAAGAA	534	
Db	131 CCTGAAGATGGCCCTTCGCCCTGAGAAGATTGACAGCTTAGTAGAGTAGAAGAG	190	
OY	535 GATGAGAGAGATGAGCGAACCTTCGCCAGGCTCCAAAGACAGCTGAAGTTGAAGCT	594	
Db	191 AATGAGAGAGATGAGCGAACCTTCGCCAGGCTCCAAAGACAGCTCAAGTTGAAGGCT	250	
OY	595 CCAGTGGACGCTGTCCATGAGAGAGAGACGCGGSCAGAGTGAACCTTGATGA	654	
Db	251 GCAGTGTGATGTCTCTGTGAGAGAGAGACGCGGSCAGAGTGAACCTTGATGA	310	
OY	655 CTTTGACAACACGGCAAGGTCACTCCGAGAGACATCAACAGCTTGTGCACACCATTCTA	714	
Db	311 CTTCGACACAAATGGCAAAAGTGAACCCGTGAGGACATTAACAGCTTGTGCATCACTCTA	370	
OY	715 TGAGGTGTGATCTCTCTGTGCACCATCCCCCAATCCAGGAAGATGCTGGCGGTAA	774	
Db	371 TGAAGTGTGATCTCTCTGTGCACCATCCCCCAATCCAGGAAGACATCTGGCGGTAA	430	
OY	775 GCTCACCCTGGCCCCCGATGGCAGCCAGAGCAAGAGAGCGTCTTGCAATCAGAGCTGA	834	
Db	431 GCTCACCATGGCTCTTGAAGAGAGCCAGAGTGAAGAGGCTCTTTTCAACATACCGA	490	
OY	835 CCTGCAGAGCGCAAGGCCCCGAGCAGAGACCAACCACTGAGGACTTCGGGACTGGGA	894	
Db	491 TTCTCAGAGGACAAAGGCCCGAGCAGACCAACCAACCCGTGAGGAGCTGCGTGGGA	550	
OY	895 GAAGAGACGAGAGCCCGGCTCAGGTTTCAGGGTGAACGCCGCTTGAGACAGTGTGCTG	954	
Db	551 GAAGAAGAGAGAGCCCGCCTCAGGTTTCAGGGTGAACGCCGCTTGAGACAGTGTGCTG	610	
OY	955 CTACCAACATTGGCTGATGAGCAATCGAGAGAGAAACCACTAATTGATCTCGCCGG	1014	
Db	611 CTATCAACCAATGGCTGATGAGCAATGAGAGAGAAACCACTAATTGATCTCGCCGG	670	
OY	1015 GATAGAAAATACTACAGTCCCAATTTGGGCGCTGCCTCCCTTCGCTGGGCCAGAGTCA	1074	
Db	671 GATAGAGAATACTACAGTCTCAGTTTGAACGGGATTCCTCTTCGCTGGGCCAGAGTCA	730	
OY	1075 ACTGCCCCCGGACCTTCCAATTCCTCATCTCGCTCCCATAGAGCCGAAAGCATCCA	1134	
Db	731 GCTGCCCCCGGATCTTCCAACCCCACTCGCTCTCGCTCCCAAGAGCAGAAAGTGCACA	790	
OY	1135 CATCCCAACCGAAAGCCCAAAGCGTGAACCCGGCTCTTCCACTTCTTGACATCCC	1194	
Db	791 CATCCCAACCGAGGGCCCAAAGGTGAACCCAGGCTCTTCCACTTCTTGACATCCC	850	

QY	1195	AATGCCCAAGGCTCTCAGAGCTTCCAGCAACGGCTCCGGGGGACACCAAGAGCGGAGAGCA	1254
Db	851	ATTTGCCAAAGGATCAAGAGCTTCCAGCAACGGCTCCGGGGCACTCAGATGGGAGCAGCA	910
QY	1255	CTTTGTGAGGTCCCCCAAAGGCCCAAGGCAAGTGTGGGTGTGGCCACGTGGCCAGAG	1314
Db	911	CTTTGTGAGGTCCCCCAAAGGCCCAAGGCAAGTGTGGGTGTGGCCACGTGGCCAGAG	970
QY	1315	GGCAAGAAACAAGCCCCCTCTGGGACCCGCCATCCCTGGGTGTGCTCCCTCCGCCACT	1374
Db	971	TGCAAGAAAGCAAGCTTCCACTGTGATCCCAACCACTATCTGTCTCCCTCTGGCCACT	1030
QY	1375	GGCTGCGAGCCGCGGCTCTCCCTCCCTAGCCCTCCGAGCCCAAGAACCAAGCA	1434
Db	1031	GGCCACAGAGCCCAAGCTCTTCTCCCACTCCCTGGACCTCTGGGGGCAAGAAACAAGCA	1090
QY	1435	CCGAGCCAGAGAGAGCCAGCAGAGGGCTGCGGGCCCTGCAAGCACTAGCTCTCAAGTG	1494
Db	1091	TGAGAGCCAGAGAGAGCCAGGCGAGCTCGGGGGCTTGCAAGGGCCCTCGCTGCAAGAG	1150
QY	1495	---CCCTGCTCGGGGGCGGAGACCTGCGGGAGCTGCGCCCTTGGTGGTGTAGAG	1551
Db	1151	CTCCACCTCATGGGGCGGAGAGAGTGAAGTGTGCTCGGTGGTGTAGAG	1210
QY	1552	CCAGGCGCGGAGCGGCTCCAGAGCATAGACACCAAGCAAGATGAATCAACCA	1611
Db	1211	CCAGGCTGGGAGCGGCTCCAGAGCAAGAACCATACCAACAGAACATCAACCA	1270
QY	1612	TTACCAACCATTTTACAGACATAGAGCCCTCCCAAGGGCCCACTGTCATATGAG	1671
Db	1271	TTACCAACCATTTTATCAGCCCTAGACCC-----CAGCAGGCTGCAAG	1319
QY	1672	GACCCCAACCCCGACACCAAGGATATTTCTATTAATTTATTAATTAATTAAT	1731
Db	1320	GACCCAGCCC--ACACCTTAAGGATTAATTTCTATTAATTTATTTATTAATTAAT	1376
QY	1732	TATTTTATTAATTAATTTATTTGTTACTCCACTTAATTTATTTAGCTAGCCATGATAGAT	1791
Db	1377	TATTTTATTAATTAATTTATTTGTTACTCCACTTAATTTATTTAGCCATGATAGAT	1436
QY	1792	CTATGGAACACAGAACTTAATTTT 1821	
Db	1437	ACATGGAACACAGAACTTAATTTT 1466	
RESULT 5			
LOCUS	AY412098	885 bp	DNA
DEFINITION	Homo sapiens NKD1 gene, VIRUTUAL TRANSCRIPT, partial sequence,		
ACCESSION	AY412098		
VERSION	AY412098.1		
KEYWORDS	GSS.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 885)		
	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
	Todd,M.A., Tenenbaum,D.M., Ciyello,D.R., Lu,F., Murphy,B.,		
	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,		
	Adams,M.D. and Cargill,M.		
	Inferring nonneutral evolution from human-chimp-mouse orthologous		
	gene trices		
TITLE	Science 302 (5652), 1960-1963 (2003)		
JOURNAL	14671302		
PUBLISHED	2 (bases 1 to 885)		
REFERENCE	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
AUTHORS	Todd,M.A., Tenenbaum,D.M., Ciyello,D.R., Lu,F., Murphy,B.,		
	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,		
	Adams,M.D. and Cargill,M.		
	Inferring nonneutral evolution from human-chimp-mouse orthologous		
	gene trices		
TITLE	Science 302 (5652), 1960-1963 (2003)		
JOURNAL	14671302		
PUBLISHED	2 (bases 1 to 885)		
REFERENCE	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,		
AUTHORS	Todd,M.A., Tenenbaum,D.M., Ciyello,D.R., Lu,F., Murphy,B.,		
	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,		
	Adams,M.D. and Cargill,M.		
	Inferring nonneutral evolution from human-chimp-mouse orthologous		
	gene trices		
TITLE	Science 302 (5652), 1960-1963 (2003)		

[illegible]

QY 1256 TTTGTAGAGTCCCCAGGCGGAGGAGAGTGTGGGTGGGC 1300
|||||
Db 841 TTGTGAGGTCCCAAGGCCAGGGCAAGAGTGTGGGTGGGC 885

RESULT 6
B0645656
LOCUS B0645656
DEFINITION ABENCCOURT_8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
5', mRNA sequence.
ACCESSION B0645656
VERSION B0645656.1 GI:21769828
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 921)
1 NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2484 row: d column: 05
High quality sequence start: 16
High quality sequence stop: 685.
Location/Qualifiers
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6285268"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_11b="NIH MGC 100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCGAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 44.5%; Score 826.8; DB 5; Length 921;
Best Local Similarity 97.6%; Pred. No. 9,8e-167;
Matches 861; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 1065 AGAAGTCAGAACTGCCCCCGGACCTTCAATCCACTGATCTCGTCCATGAGCCGG 1124
|||||
Db 263 AGAAGTCAGAACTGCCCCCGGACCTTCAATCCACTGATCTCGTCCATGAGCCGG 322

QY 1125 AAGCATTCACATCCCAACCGAAGCCCGAGCGGTGAACCCGCGCTCTTCACTTCC 1184
|||||
Db 323 AAGCATTCACATCCCAACCGAAGCCCGAAGCGGTGAACCCGCGCTCTTCACTTCC 382

QY 1185 TTGACACCCCAATTCGCCAAGTCTCAGAGTCCAGCAACGGCTTCGGGGGACCCAGAGACG 1244
|||||
Db 383 TTGACACCCCAATTCGCCAAGTCTCAGAGTCCAGCAACGGCTTCGGGGGACCCAGAGACG 442

QY 1245 GAGCAAGCACTTTGAGAGTTCGCCCAAGGCCCGAGGGCAAGAGTGTGAGTGGAGCAGC 1304
|||||
Db 443 GAGCAAGCACTTTGAGAGTTCGCCCAAGGCCCGAGGGCAAGAGTGTGAGTGGAGCAGC 502

QY 1305 TGCCAGAGGGGCAAGAAACAAAGCCCTCTGAGAACCCGCAATCCCTGCGGTCTCCCT 1364
|||||
Db 503 TGCCAGAGGGGCAAGAAACAAAGCCCTCTGAGAACCCGCAATCCCTGCGGTCTCCCT 562

QY 1365 CCGGCCACTGTGCTGCCAGCCCGGCTCTCTCCCTTCTTGAAGCCCTTCTGGGCAAGAGA 1424
|||||
Db 563 CCGGCCACTGTGCTGCCAGCCCGGCTCTCTCCCTTCTTGAAGCCCTTCTGGGCAAGAGA 622

QY 1425 AGCAACAGCACCGAGCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1484
|||||
Db 623 AGCAACAGCACCGAGCCAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 682

QY 1485 CTTGAGGTGGCCCTGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1544
|||||
Db 683 CTTGAGGTGGCCCTGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 742

QY 1545 ATGAGAGCCAGGCGCGGAGCCGCTCCAGAGCATGAGACCAACCAACCATGAAATATC 1604
|||||
Db 743 ATGAGAGCCAGGCGCGGAGCCGCTCCAGAGCATGAGACCAACCAACCATGAAATATC 802

QY 1605 ACACCATTCACCAACCTTACCAAGACAT--GAGCCCTCCCGAGGGGCCCACTTG- 1661
|||||
Db 803 ACACCATTCACCAACCTTACCAAGACATGAAACCCCTTCCCGAGGGGCCCACTTG 862

QY 1662 -CCATATGAAAGACCCCAAGCCCGGAGCAACCAAGCATTAAT 1702
|||||
Db 863 CCATATGAAAGACCCCAAGCCCGGAGCAACCAAGCATTAAT 904

RESULT 7
B0653673
LOCUS B0653673
DEFINITION ABENCCOURT_8419128 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284223
5', mRNA sequence.
ACCESSION B0653673
VERSION B0653673.1 GI:21777845
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 931)
1 NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2481 row: h column: 16
High quality sequence stop: 656.

REFERENCE 1 (bases 1 to 953)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@ds-riemail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: L16M2784 row: d column: 18
 High quality sequence stop: 627.
 Location/Qualifiers
 source
 1..953
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6580002"
 /tissue_type="teratocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_109"
 /note="Organ: ovary; Vector: pOT87; Site 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 40.8%; Score 757.6; DB 5; Length 953;
 Best Local Similarity 97.5%; Pred. No. 7,1e-152;
 Matches 833; Conservative 0; Mismatches 14; Indels 7; Gaps 6;

QY 66 AGTGGGCGGGGAGAGCGGCGGAGCGGCTCCGCGGCTCCGCTCGG 125
 |||||
 DB 66 AGTGGGCGGGGAGAGCGGCGGAGCGGCTCCGCGGCTCCGCTCGG 125
 |||||
 QY 126 CTCGGGGGCTGCTTCGGGAGAGAGAGCCAGAGAGCGCCGCGCGGCGG 185
 |||||
 DB 126 CTCGGGGGCTGCTTCGGGAGAGAGAGCCAGAGAGCGCCGCGCGGCGG 185
 |||||
 QY 186 CATGGCTTAAAGAGCGTCCCGCGCGCGAGCCACAGATGGGAAACTTCACTCAAGC 245
 |||||
 DB 186 CATGGCTTAAAGAGCGTCCCGCGCGCGAGCCACAGATGGGAAACTTCACTCAAGC 245
 |||||
 QY 246 CGGCGCGCGGTGCAAGCGAGAGAGCGCGGAAAGGTGACAGTTGGCGGTGAGCGCTG 305
 |||||
 DB 246 CGGCGCGCGGTGCAAGCGAGAGAGCGCGGAAAGGTGACAGTTGGCGGTGAGCGCTG 305
 |||||
 QY 306 CTTGGGCTCGGAAGGAGCATCGAGAGTGTATCGGAGACAGCGCTGCCGCGGCTGCT 365
 |||||
 DB 306 CTTGGGCTCGGAAGGAGCATCGAGAGTGTATCGGAGACAGCGCTGCCGCGGCTGCT 365
 |||||
 QY 366 CGGAGACCCCGACAGCTCGGTGGCGGCGGACCATATGGCCGGAAGACCCGGGAGCTCGTGG 425
 |||||
 DB 366 CGGAGACCCCGACAGCTCGGTGGCGGCGGACCATATGGCCGGAAGACCCGGGAGCTCGTGG 425
 |||||
 QY 426 GCGAGGTGTGAGAGACACGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
 |||||
 DB 426 GCGAGGTGTGAGAGACACGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
 |||||
 QY 486 CCTGCTCTCTGAG 545
 |||||
 DB 486 CCTGCTCTCTGAG 545
 |||||
 QY 546 TGAAGCAACCTCTGCCAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
 |||||
 DB 546 TGAAGCAACCTCTGCCAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
 |||||

QY 606 TGTTCATGAG 665
 |||||
 DB 606 TGTTCATGAG 665
 |||||
 QY 666 ACGGCAAGGTACCCGAGAGAGATCAACAGCTTGTGCAACACATATATAGGTGTGG 725
 |||||
 DB 666 ACGGCAAGGTACCCGAGAGAGATCAACAGCTTGTGCAACACATATATAGGTGTGG 725
 |||||
 QY 726 ACTTCTTGTCAACCACTCCCAACATCCAGCAAGATGCTG-GGGTAAAGCTCAACGCTG 784
 |||||
 DB 726 ACTTCTTGTCAACCACTCCCAACATCCAGCAAGATGCTGCGGGTAAAGCTCAACGCTG 785
 |||||
 QY 785 G-CCCCGATGGAG 841
 |||||
 DB 786 GCCCCGATGGAG 845
 |||||
 QY 842 AGCGCAAGGCCCGAG 898
 |||||
 DB 846 AGCGCAAGGCCCGAG 905
 |||||
 QY 899 AAGCAG 912
 |||||
 DB 906 AAGCAG 919
 |||||

RESULT 12
 AY412099 885 bp DNA linear GSS 16-DEC-2003
 LOCUS Pan troglodytes NKD1 gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 AY412099
 VERSION AY412099.1 GI:39768064
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1 (bases 1 to 885)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Snihsky,D.J.,
 Adams,M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 14671302
 2 (bases 1 to 885)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Snihsky,D.J.,
 Adams,M.D. and Cargill,M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 Location/Qualifiers
 source
 1..885
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>885
 /gene="NKD1"
 /locus_tag="HGM4444"

ORIGIN
 Query Match 40.3%; Score 748.8; DB 9; Length 885;
 Best Local Similarity 84.7%; Pred. No. 5,4e-150;
 Matches 750; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 416 GAGCTCGTGGGCAAGGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGAGAGAG 475
 |||||

[illegible]

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgabbs-remail.nih.gov
           Tissue Procurement: CGAP (Stanford)
           CDNA Library Preparation: Rubin Laboratory
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLNL)
           DNA sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNLNL at:
           http://image.lnl.nih.gov
           Plate: LILCM2447 row: b column: 19
           High quality sequence stop: 583.
           Location/Qualifiers
               1..966
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:62711026"
               /tissue_type="hepatocellular carcinoma, cell line"
               /lab_host="DH10B (phage-resistant)"
               /clone_id="NIH_MGC_100"
               /note="Organ: liver; Vector: pOTB1; Site 1: XhoI; Site 2:
               EcoRI, cDNA made by oligo-dT priming. Directionally cloned
               into EcoRI/XhoI sites using the following 5' adaptor:
               GGCCGAGGAG(G). Size selected >500bp for average insert size
               1.8kb. Library constructed by Ling Hong in the laboratory
               of Gerald M. Rubin (University of California, Berkeley)
               using ZAP-cDNA synthesis kit (Stratagene) and Superscript
               II RT (Life Technologies). Note: this is a NIH_MGC
               library."

```

Query Match	38.3%	Score 712.6	DB 5	Length 966
Best Local Similarity	94.6%	Pred. No. 3.3e-142		
Matches	759	Conservative	0	Mismatches 40
			Indele	3
			Gaps	2
QY	825	ATCAGGCTGACCTTSCAAGCCGACAGGCCCCCGACGACAGACCAGCCCACTGAGACCTGC	884	
DB	3	ATCAGGCTGACCTGACAGCGCAGAGCCCGACGACAGACCAGCCCACTGAGACCTGC	62	
QY	885	GGAGCTGGAGAAAGAGCAGCGAGCCCGCTCAGGTTCCAGGGTGAACAGCCGCTTGAGAC	944	
DB	63	GGAGCTGGAGAAAGAGCAGCGAGCCCGCTCAGGTTCCAGGGTGAACAGCGCTTGAGAC	122	
QY	945	AGCTGGGTGTACCAACATTTGGGCTAGTGAACATCGAGAGAGAAACCACTACTTAG	1009	
DB	123	AGCTGGGTGTACCAACATTTGGGCTAGTGAACATCGAGAGAGAAACCACTACTTAG	182	
QY	1005	ATCTCGCCGGGATAGAAAACCTACACGTCCTCCAAATTGGGGCTGGCTCCCTTCGGTGGCC	1064	
DB	183	ATCTCGCCGGGATAGAAAACCTACACGTCCTCCAAATTGGGGCTGGCTCCCTTCGGTGGCC	242	
QY	1065	AGAGTCGAACTGCCCCCCCCGCACTTCCATCCACTCGATCTCGTCCCATGAGCCGG	1124	
DB	243	AGAGTCGAACTGCCCCCCCCGCACTTCCATCCACTCGATCTCGTCCCATGAGCCGG	302	
QY	1125	AAGCCATCCATCCACACCGAAAGCCCCAAGGGGTGAGACCGGCGCTCTTCCACTTCC	1188	
DB	303	AAGCCATCCATCCACACCGAAAGCCCCAAGGGGTGAGACCGGCGCTCTTCCACTTCC	362	
QY	1185	TTGACACCCCAATGCGCAAGGTTCTCAGAGCTTCAGCAACCGGCTCCGGGGACCCAGAGC	1244	
DB	363	TTGACACCCCAATGCGCAAGGTTCTCAGAGCTTCAGCAACCGGCTCCGGGGACCCAGAGC	422	
QY	1245	GGAGCAAGCACTTTGTGAGTCCCCCAAGGCCCAAGGGCAAGTGTGGGTGTGGGCCAGC	1304	
DB	423	GGAGCAAGCACTTTGTGAGTCCCCCAAGGGGCCAAGGGCAAGTGTGGGTGTGGGCCAGC	482	
QY	1305	TGCGCAGAGGGGCAAGAAACAAGCCCTCTTGAGACCGCGCATCTCTGCGTGTCCCTT	1364	
DB	483	TGCGCAGAGGGGCAAGAAACAAGCCCTCTTGAGACCGCGCATCTCTGCGTGTCCCTT	542	

QY 1365 CCGGCCACCTGGCTGCGACGCCGCGCTCCCTCCCTAGCCGCCCTCGGGGACAGA 1424
DB 543 CCGGCCACCTGGCTGCGACGCCGCGCTCCCTCCCTAGCCGCCCTCGGGGACAGA 602
QY 1425 AGCACAAGCAGCCGAGCCAGAGAGAGCCAGAGAGGCTGCGGGGCTTCAGGACCACTGG 1484
DB 603 AGCACAAGCAGCCGAGCCAGAGAGAGCCAGAGAGGCTGCGGGGCTTCAGGACCACTGG 662
QY 1485 CCGCAGGTGGCCCTCTCTCTCTGGGGGGGAGACACTG--CGGAGCTGCCCTGTGTGT 1542
DB 663 CCGCAGGTGGCCCTCTCTCTGGGGGGGAGACACTGCTTGGCCGCTCGGGGGGC 722
QY 1543 GTATGAGAGCAGGCGCGGAGCC--GGTCCAGAGACATGAGACCAACCACTATGAC 1601
DB 723 GTATGAGAGCAGGCTGGGCTCCGGGCTTCAGGAGCTTCAACCCCACTAGAGGAG 782
QY 1602 ATGACACCATTAACCACTT 1623
DB 783 GTCCCCCGCTTCCCGCGCT 804

RESULT 14
B0652087 996 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8207495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283110
DEFINITION 5', mRNA sequence.
ACCESSION B0652087
VERSION B0652087.1 GI:21776259
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM2478 row: j column: 07
High quality sequence stop: 604.
Location/Qualifiers
1..996
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6283110"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

QY 822 TCATATCAGGTGACCTGACAGCCGACAGGCCCGACAGAGACCAAGCCCACTGAGAGAC 881

DB 1 TCATATCAGGTGACCTGACAGCCGACAGGCCCGACAGAGACCAAGCCCACTGAGAGAC 60
QY 882 TCGGAGCTGGGAGAGAGAGACGAGCCCGCTCAGGTTCCAGGGTACAGCCGCTGG 941
DB 61 TCGGAGCTGGGAGAGAGAGACGAGCCCGCTCAGGTTCCAGGGTACAGCCGCTGG 120
QY 942 AGCAGTCTGGCTCTCAACCATTTGGTATGAGAAACATCGAGAGAGAAACCACTACT 1001
DB 121 AGCAGTCTGGCTCTCAACCATTTGGTATGAGAAACATCGAGAGAGAAACCACTACT 180
QY 1002 TAGATCTCGCCGGATAGAAAATTAACAGTCCCAATTTGGGCTGCTCCCTTCGTGG 1061
DB 181 TAGATCTCGCCGGATAGAAAATTAACAGTCCCAATTTGGGCTGCTCCCTTCGTGG 240
QY 1062 CCCAGAGTCAAGATGCCCCCGGACCTCCATCCACTCGATCTCGTCCCATGAGC 1121
DB 241 CCCAGAGTCAAGATGCCCCCGGACCTCCATCCACTCGATCTCGTCCCATGAGC 300
QY 1122 CCGAAGCCATCCACATCCACACCGAAGCCCGAAGCCCGGCTCCCTTCACCT 1181
DB 301 CCGAAGCCATCCACATCCACACCGAAGCCCGAAGCCCGGCTCCCTTCACCT 360
QY 1182 TCCTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTCCGGGAGACCCAGG 1241
DB 361 TCCTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTCCGGGAGACCCAGG 420
QY 1242 ACGGAGCAAGACCTTTGTAGAGTCCCGCAAGGCCCAAGAGTGTGTGTGTGGCC 1301
DB 421 ACGGAGCAAGACCTTTGTAGAGTCCCGCAAGGCCCAAGAGTGTGTGTGTGGCC 480
QY 1302 ACGTGGCCAGAGGGGAGAAACAAGCCCTCTGGAGCCCGCATCCCTGGGTGTGCC 1361
DB 481 ACGTGGCCAGAGGGGAGAAACAAGCCCTCTGGAGCCCGCATCCCTGGGTGTGCC 540
QY 1362 CTTCCGCCCACTGGTGGCAGGCCGCGCTCTCCCTCCCTAGGCCCTTCGGGACCA 1421
DB 541 CTTCCGCCCACTGGTGGCAGGCCGCGCTCTCCCTCCCTAGGCCCTTCGGGACCA 600
QY 1422 AGAAGCAACAAGCAGCCAGCCAGAGAGAGCCAGAGGCTGCGGGCTTCAGAGCACAC 1481
DB 601 AGAAGCAACAAGCAGCCAGCCAGAGAGAGCCAGAGGCTGCGGGCTTCAGAGCACAC 660
QY 1482 TGGCCTCAGTGGCCCTCTCTCTGGGGGGGAGACCTGGGGAGCTGCCCGCTT--GG 1538
DB 661 TGGCCTCAGTGGCCCTCTCTCTGGGGGGGAGACCTGGGGAGCTGCCCGCTTGGCGG 720
QY 1539 TGGTGTATGAGAGCCAGGCGGGGAGCCGCTCCAGAGACATGAGACCAACCAACC 1595
DB 721 GGGTATGAGAGACCCAGGCCCGGCAACCGGTCCCAAAAATTGACCCCTTCACC 777

RESULT 15
B0646371 1008 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8493065 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6300188
DEFINITION 5', mRNA sequence.
ACCESSION B0646371
VERSION B0646371.1 GI:21770543
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1008)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2514 row: a column: 21
 High quality sequence start: 17
 High quality sequence stop: 553.
 Location/Qualifiers

FEATURES

1..1008
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6300188"
 /issue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NIH MGC 100"
 /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the Laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

ORIGIN

Query Match 38.0%; Score 706; DB 5; Length 1008;
 Best Local Similarity 92.5%; Pred. No. 8.5e-141;
 Matches 835; Conservative 0; Mismatches 46; Indels 22; Gaps 8;

```

QY      825 ATCAGGCTGACTGCAAGAGCGCAAGGCCCGACAGACCAAGCCACTGAGACCTGC 884
      |||
Db      23 ATCAGGCTGACTGCAAGAGCGCAAGGCCCGACAGACCAAGCCACTGAGACCTGC 82

QY      885 GGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
      |||
Db      83 GGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 142

QY      945 AGTCTGGCTGCTACCACTTGGCTAGATGAGAACTGAGAGAGAGAGAGAGAGAGAGAG 1004
      |||
Db      143 AGTCTGGCTGCTACCACTTGGCTAGATGAGAACTGAGAGAGAGAGAGAGAGAGAGAG 202

QY      1005 ATCTGCCCGGATGAGAACTACAGCTCCCAATTTGGGCTGCTCCCTTCCTGAGCC 1064
      |||
Db      203 ATCTGCCCGGATGAGAACTACAGCTCCCAATTTGGGCTGCTCCCTTCCTGAGCC 262

QY      1065 AGAAGTCAGAACTGCCCCCGGACCTCCAAATCCCACTGATTCGCTCCATGAGCCGG 1124
      |||
Db      263 AGAAGTCAGAACTGCCCCCGGACCTCCCAATCCCACTGATTCGCTCCATGAGCCGG 322

QY      1125 AAGCCATCCACATCCCAACCGAAAGCCCAAGGCGTGAACCGGCTCTTCCACTTCC 1184
      |||
Db      323 AAGCCATCCACATCCCAACCGAAAGCCCAAGGCGTGAACCGGCTCTTCCACTTCC 382

QY      1185 TTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTCGGGGCAACCAGAGCG 1244
      |||
Db      383 TTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTCGGGGCAACCAGAGCG 442

QY      1245 GGAGCAAGCACTTTGAGAGTCCCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAAG 1304
      |||
Db      443 GGAGCAAGCACTTTGAGAGTCCCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAAG 502

QY      1305 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGAGAACCCGCCAT-CCCTGCGGTGTCCCC 1363
      |||
Db      503 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGAGAACCCGCCATCCCTGCGGTGTCCCC 562

QY      1364 TCGGCCACCTGGCTGCCAGCCGCGCTCTCTCCCTCCCTAGCCCCCTCGGGCACAAG 1423
      |||
Db      563 TCGGCCACCTGGCTGCCAGCCGCGCTCTCTCCCTCCCTAGCCCCCTCGGGCACAAG 622

QY      1424 AAGCAACAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1483
      |||

```

```

Db      623 AAGCACAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 682
      |||
QY      1484 GCCTCAGGTGGCCCTG-TCCTGGGGCGGAGACAC-TCGGGAGCTGCCGC---CTTG 1537
      |||
Db      683 GCCTCAGGTGGCCCTGTTCTCTGGGGCGGAGACACCTTGGGGAGCTGCCGCCGCTTGGG 742
      |||
QY      1538 GTGTGTATAGAGCCAGAGCC--GGGAGCCGCTCCAGAGACATGAGACACCA---C 1591
      |||
Db      743 GGGGGGTATGAAAGCCCGGNCGGGGCAGCGGGGTCCAGAAGACATGAGACACCAAC 802
      |||
QY      1592 CACCATGAACATACCAACCATTTACCAACCTTACCAAGATAGAGCC---TCCCC 1647
      |||
Db      803 CACCATGAACATACCAACCATTTACCAACCTTACCAAGATAGAGCC---TCCCC 862
      |||
QY      1648 AGGGCCCAACCTGCATA----TGAAGAGCCCAACCCCGGACACCAAGGCAATTAT 1702
      |||
Db      863 AGGGGCAAAACCGGGGCAAAATGAAAGAAACCAACCCCGGAAACCCCAAGGTTTATT 922
      |||
QY      1703 ATT 1705
      |||
Db      923 AAT 925
      |||

```

Search completed: December 29, 2004, 20:34:18
 Job time : 8801 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 03:23:44 ; Search time 141 Seconds
(without alignments)
2369.296 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 2499
Sequence: 1 MGJLHSHKPAVCKRRSPBG.....RHHHHHHHHHHHHYHPTQT 470

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-Q/cgnt2.1/USPTO.spool.P/US0993966/runat.29122004.130759.9030/app.query.fasta.1.647
-DB=Issued Patents NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0993966 @CGN 1.128 @runat.29122004.130759.9030 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued Patents NA.*
1: /cgnt2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgnt2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgnt2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgnt2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgnt2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgnt2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2493	99.8	1438	US-09-506-066E-5	Sequence 5, Appl1
2	2174.5	87.0	1731	US-09-506-066E-3	Sequence 3, Appl1
3	783	31.3	1285	US-09-506-066E-7	Sequence 7, Appl1
4	583.5	23.3	1307	US-09-506-066E-9	Sequence 9, Appl1
5	215.5	8.6	4954	US-09-506-066E-1	Sequence 1, Appl1
6	160.5	6.4	2063	US-09-270-767-14537	Sequence 14537, A
7	153	6.1	1449	US-09-252-991A-15134	Sequence 15134, A
8	153	6.1	2226	US-09-252-991A-14877	Sequence 14877, A
9	152	6.1	2268	US-09-252-991A-14487	Sequence 14487, A
10	152	6.1	2433	US-08-999-689A-2	Sequence 2, Appl1
11	152	6.1	2433	US-09-944-807-3	Sequence 3, Appl1
12	149.5	6.0	4403765	US-09-103-840A-2	Sequence 2, Appl1

13	149.5	6.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
14	146.5	5.9	2576	1	US-08-471-033-35	Sequence 35, Appl1
15	146.5	5.9	2576	2	US-08-471-044-35	Sequence 35, Appl1
16	146.5	5.9	2576	2	US-08-463-483A-35	Sequence 35, Appl1
17	146.5	5.9	2576	2	US-08-471-046A-35	Sequence 35, Appl1
18	146.5	5.9	2576	2	US-08-470-566B-35	Sequence 35, Appl1
19	146.5	5.9	2576	2	US-08-469-334-35	Sequence 35, Appl1
20	146.5	5.9	2576	2	US-09-300-529-35	Sequence 35, Appl1
21	146.5	5.9	2655	1	US-08-471-033-17	Sequence 17, Appl1
22	146.5	5.9	2655	1	US-08-471-033-26	Sequence 26, Appl1
23	146.5	5.9	2655	2	US-08-471-044-17	Sequence 17, Appl1
24	146.5	5.9	2655	2	US-08-471-044-26	Sequence 26, Appl1
25	146.5	5.9	2655	2	US-08-463-483A-17	Sequence 17, Appl1
26	146.5	5.9	2655	2	US-08-463-483A-26	Sequence 26, Appl1
27	146.5	5.9	2655	2	US-08-471-046A-17	Sequence 17, Appl1
28	146.5	5.9	2655	2	US-08-471-046A-26	Sequence 26, Appl1
29	146.5	5.9	2655	2	US-08-470-566B-17	Sequence 17, Appl1
30	146.5	5.9	2655	2	US-08-470-566B-26	Sequence 26, Appl1
31	146.5	5.9	2655	2	US-08-469-334-17	Sequence 17, Appl1
32	146.5	5.9	2655	2	US-08-469-334-26	Sequence 26, Appl1
33	146.5	5.9	2655	2	US-09-300-529-17	Sequence 17, Appl1
34	146.5	5.9	2655	3	US-09-300-529-26	Sequence 26, Appl1
35	146.5	5.9	4031	1	US-08-471-033-49	Sequence 49, Appl1
36	146.5	5.9	4031	2	US-08-471-044-49	Sequence 49, Appl1
37	146.5	5.9	4031	2	US-08-463-483A-49	Sequence 49, Appl1
38	146.5	5.9	4031	2	US-08-471-046A-49	Sequence 49, Appl1
39	146.5	5.9	4031	2	US-08-470-566B-49	Sequence 49, Appl1
40	146.5	5.9	4031	2	US-08-469-334-49	Sequence 49, Appl1
41	146.5	5.9	4031	3	US-09-300-529-49	Sequence 49, Appl1
42	146.5	5.8	5194	1	US-09-599-652-1	Sequence 1, Appl1
43	146.5	5.8	5194	2	US-08-642-846-1	Sequence 1, Appl1
44	146.5	5.8	5194	3	US-09-264-604-1	Sequence 1, Appl1
45	146.5	5.8	5194	4	US-10-002-389-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
TITLE OF INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIORITY FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8) ..(1418)
OTHER INFORMATION: Nkd1 coding sequence
US-09-506-066E-5
Alignment Scores:
Pred. No.: 8.39E-206
Score: 2493.00
Percent Similarity: 99.79%
Best Local Similarity: 99.79%
Query Match: 99.76%
DB: 4
Gaps: 0
US-09-993-966-7 (1-470) x US-09-506-066E-5 (1-1438)
Length: 1438
Matches: 469
Conservative: 0
Mismatch: 1
Indels: 0

```

QY 1 MercGlyLeuHisSerIysProAlaValCysIleAspArgGluSerProGluGly 20
Db 8 ATGGGGAACTTCACTCAAGCCGGCCGGCTGTGACGACGAGGAGCCGGAAGGT 67
QY 21 AspSerPheAlaValSerAlaAlaTrpAlaArgIleGlyIleGluTrpIleGlyArg 40
Db 68 GACAGCTTCGGCCGTGACGGCTGGCTGGAGGAGCATCGAGAGTGGATGGGAGA 127
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db 128 CAGCGCTTCGGCCGGCTGTCTCCGGGACCCCGACAGCTGCGGTGGGGGACCATGAGC 187
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCCGGAGCTGTGGGCGACGTGTGAGACACGCTCAGCGAGAGAGAG 247
QY 81 AspArgPheArgLeuGluValAlaLeuProProGluIleuArgIleuGlySerGly 100
Db 248 GACGACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAAAGCTGACGGGCTGGGACGGGA 307
QY 101 ArgGluIleuValMetGluArgValSerGluProCysProGlySerIleuValLeuLys 120
Db 308 GATGAGAAAGAGAGAGAGAGTACGCAACCTGCGCAGGCTCCAGAGACAGCTGAG 367
QY 121 PheGluGluLeuGluGlnCysAspValSerMetGluIleuAspSerArgGlnGluTrpThrPhe 140
Db 368 TTTCAGAGCTCCAGTGTGAGGTTCATGTGAGAGAGACCCGCGAGAGTGAACCTTC 427
QY 141 ThrLeuTrpAspPheAspAsnAsnGlyIleValThrArgIleuAspIleThrSerLeuLeu 160
Db 428 ACCGTATGACTTTGACAAACAGGAGTCAACCGAGAGGACATCACAGCTTCGTCG 487
QY 161 HisThrIleTrpGluValAlaAspSerSerValAsnHisSerProThrSerSerIleMet 180
Db 488 CACACCATCATATGAGGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATG 547
QY 181 LeuArgValIleuSerThrValAlaProAspGlySerGlnSerIleuValSerValLeuVal 200
Db 548 CTGGGGGTAAAGCTCACCGTGGCCCGATGGCGAGCCAGACAGAGAGAGGCTTCCTGTC 607
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgIleuArgIleuThrIleuAspLeu 220
Db 608 AATCAGGCTGACTCGACAGAGCGCAAGGCCCGACAGAGACCAAGCCCATGAGAGACTG 667
QY 221 ArgSerTrpGluIleuValGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGGAGCTGGAGAGAGAGCGAGCGCCCGCTCAGGTTCGAGGTGACAGCGGCTGGAG 727
QY 241 GluSerGlyCysTrpHisAspValAspGluAsnIleGluArgArgAsnHisTrpLeu 260
Db 728 CAGCTGGCTCTTACCACTTGGCTGATGAGACATCGAGAGAGAAACCACTACTTA 787
QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db 788 GATCTCGCGGGATAGAAACTACACGTCCTCATTTGGGCTCGCTCCCTTCGTCGAGC 847
QY 281 GluIleuSerGluLeuProArgTrpThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 848 CAGAGGTCAAGAACTGCCCCCGACCTTCATATCCATCTCATCTCCCTCCAGAGACCG 907
QY 301 GluAlaIleHisIleProHisArgIleProGlnGlyValAspProAlaSerPheHisPhe 320
Db 908 GAAACCATCCACATCCACACCGAAAGCCCAAGGCTGGACCCGGCTCTTCACCTTC 967
QY 321 LeuAspThrProIleAlaIleValSerGluLeuGlnArgLeuArgIleuThrGlnAsp 340
Db 968 CTTCACACCCCAATCGCAAGTCTCAGAGCTCCAGAGAGCGCTCCGGGACCTCAGAGAC 1027
QY 341 GlySerIleHisPheValArgSerProIleValArgIleGlySerValGlyValGlyHis 360
Db 1028 GGGGACAGGACTTGTGTGAGGTCCCGCAAGGCGCAAGGAGAGTGTGGGTGGGAC 1087
QY 361 ValAlaArgGlyAlaArgAsnIleProProLeuGluProAlaIleProAlaValSerPro 380

```

```

Db 1088 GTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGACCCGCATCTCCGTGGGTGCCCC 1147
QY 381 SerAlaHisLeuAlaAspProAlaLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1148 TCCGCCACCTGTGCTGCACGCCCGGCTCTCCCTCCCTTACGCCCTCCCGGAGCACAG 1207
QY 401 LysHisIleHisArgAlaIleArgGluSerGlnGluGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1208 AAGCACAGACCGAGCGCAAGAGAGACACAGAGGCTGCGGGGCTTCGACAGCACCTG 1267
QY 421 AlaSerGlyIleProValIleuGlyArgGluHisLeuArgGluLeuProAlaVal 440
Db 1268 GCCTCAGGTGGCCCTGTCTCGGGGCGGAGACCTGGGGAGCTGCCGCTGTGTGTG 1327
QY 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisIleHisIleGluHis 460
Db 1328 TATGAGAGCAGGCGCGGCGCAGCGCTCCAGAGACATAGACACCAACCATGAAAT 1387
QY 461 HisHisIleThrHisPheIleArgIleThr 470
Db 1388 CACCACCATTAACAACCTTCTACAGACA 1417

```

RESULT 2
US-09-506-066E-3
: Sequence 3, Application US/09506066E
: Patent No. 6630323
: GENERAL INFORMATION:
: APPLICANT: Scott, Matthew
: APPLICANT: Wharton, Keith
: APPLICANT: Zeng, Wenlin
: TITLE OR INVENTION: Naked Cuticle Genes and their Uses
: FILE REFERENCE: STRA-121
: CURRENT APPLICATION NUMBER: US/09/506,066E
: PRIOR FILING DATE: 2000-02-17
: PRIOR FILING DATE: 1999-02-17
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 1731
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (140)...(1553)
: OTHER INFORMATION: Nkd1 coding sequence
US-09-506-066E-3

Alignment Scores:
Score: 3,46e-178 Length: 1731
Percent Similarity: 2174.50 Matches: 407
Best Local Similarity: 90.64% Conservative: 19
Query Match: 86.60% Mismatches: 43
DB: 87.01% Indels: 1
Gaps: 1

US-09-993-966-7 (1-470) x US-09-506-066E-3 (1-1731)

```

QY 1 MercGlyLeuHisSerIysProAlaValCysIleAspArgGluSerProGluGly 20
Db 140 ATGGGGAACTTCACTCAAGCCGGCCGGCTGTGACGACGAGGAGCCGGAAGGT 159
QY 21 AspSerPheAlaValSerAlaAlaTrpAlaArgIleGlyIleGluTrpIleGlyArg 40
Db 200 GACAGCTTTCGGCTGGAAGTGGCCCTGCTCTGAGAAAGCTTCAAGAGTGGATCGGAGG 259
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db 260 CAGCGCTTCGGCCGGCTGTCTCCGGGACCCCGACAGCTGCGGTGGGACGACTGGGT 319
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80

```



```

Db      320 CGAGGCACTGGGAACTCTGGGTGACACTTCTAGAGAGGCTCTCGGTGAGAGAGAG 379
Qy      81 AspAspPheArgLeuGluValAlaLeuProProGluIuYrThrAspGlyLeuGlySerGly 100
Db      380 GACGACTTCCCCCTTGAAGAGTGGCCCTGCGCTGAGAAAGATGACAGAGCTGATGATGAGA 439
Qy      101 ArgGluIuYrMetGluArgValSerGluProCySerProGlySerGlyValGluLeuLys 120
Db      440 GATGAGAGAGAGAAATGAGAGACTGAGGAACTTGCGCCAGCCCTCCAGAAACCACTCAG 499
Qy      121 PheGluIuLeuGlnCySerPheValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db      500 TTTCAGAGAGCTACAGTGTGATGTCTCTGTGAGAGAGAGAGAGCCGCGAGAGTGAATTTTC 559
Qy      141 ThrLeuTrpAspPheAspAsnAnglyValValThrArgGluAspIleThrSerLeuLeu 160
Db      560 ACTCTATATGACTTCCACCAACATGGCAAGTGAACCGTGAGACATTAACAGCTTGCTG 619
Qy      161 HisThrIleTrpGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db      620 CATACCATCTATGAGAGTGTGACTCTCTGTGAAACCATTCCTCCACATCAAGCAAGACA 679
Qy      181 LeuArgValIuYrLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
Db      680 CTGGGGGTGAAGCTCACCTGCTCTCTGACGGAGCCAGAGTAAAGAGAGCTCTTTTC 739
Qy      201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
Db      740 AACCATACCGATCTGACAGAGACAAAGGCCCGAGACAGACAAACCGGTGAGAGAGCTG 799
Qy      221 ArgSerTrpGluIuYrGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db      800 CGTGCTGGAGAGAAAGACAGCCGACCTCAGGTTCACAGGTGACAGCCACCTTGAG 859
Qy      241 GlnSerGlyCySerTrpHisSerValAspGluAsnIleGluArgAsnHisTrpLeu 260
Db      860 CAGCGACACTGCTTCCACCATCTGCTGTGATGAGAACATTTGAGAGAGAAACCATTA 919
Qy      261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db      920 GACCTGGCGGGAGATGAGAACTACACGCTCAGTTTGACGGGATCCCTTCGCTG 979
Qy      281 GlnLysSerGluLeuProProArgTrpThrSerAspProThrArgSerArgSerHisGluPro 300
Db      980 CAGAAAGTCAAGAGTCCCTCGAATCTCCAAACCCACTGCTCTCCCTCCACAGACCA 1039
Qy      301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db      1040 GAACTCCCAATCCACACCGGAGGCCCCAGAGTGTGAGACCAAGGCTCTTCCACTC 1099
Qy      321 LeuAspTrpProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
Db      1100 CTTCGACACCCCATTTGCGCAAGGATCAGAGCTCCAGCAAGCGCTCCGGGGACATCAGAT 1159
Qy      341 GlySerTrpHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db      1160 GGGAGCAGAGACTTTTGAGGTTCCTCCCAAGGCCCAAGGCAAGATGGGTATGGGCAAC 1219
Qy      361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db      1220 GGGGCGAGAGGTGCAAGAGCAAGCTTCACTGTATCCACCACTCATCTCTTCCCTC 1279
Qy      381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1280 TTGGCCCATCTGGGCAACAGCCAGCCCTTCTCCACACCTGCGACCCCTGGGGCAAG 1339
Qy      401 LysHisLysHisArgAlaLysGlnSerGlnGlnLysArgGlyLeuGlnAlaProLeu 420
Db      1340 AAACACAAAGATCGAGCCAAAGAGCCAGGCGAGCTCCGGGGCTGCAAGGCCCTTCG 1399
Qy      421 AlaSerGlyGly---ProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 439
Db      1400 GCTGCAAGAGGCTCCACCTCATGAGGCGGGAGAGAGTGAAGAGCTGCTGCTGCTGATG 1459

```

```

Qy      440 ValTrpGluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisGlu 459
Db      1460 GTGTACGAGAGCCAGGCTTACAGGCCCTGCAGAGACAGAAACCATCAACCAAGAA 1519
Qy      460 HisHisHisHisTrpHisHisPheTrpGln 469
Db      1520 CATCACCAACATTTATCACCACTTATCAG 1549

RESULT 3
US-09-506-066E-7
: Sequence 7, Application us/09506066E
: Patent No. 6630923
: GENERAL INFORMATION:
: APPLICANT: Scott, Matthew
: APPLICANT: Wharton, Keith
: APPLICANT: Zeng, Wenlin
: TITLE OF INVENTION: Naked Cyticle Genes and their Uses
: FILE REFERENCE: STAN-121
: CURRENT APPLICATION NUMBER: US/09/506,066E
: PRIOR FILING DATE: 2000-02-17
: PRIOR FILING DATE: 1999-02-17
: NUMBER OF SEQ. ID NOS: 15
: SOFTWARE: PaedSeq for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1285
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (2)...(1282)
: OTHER INFORMATION: Nkd2 coding sequence
US-09-506-066E-7

Alignment Scores:
Pred. No.: 2,9e-58 Length: 1285
Score: 783.00 Matches: 180
Percent Similarity: 56.86% Conservative: 52
Best Local Similarity: 44.12% Mismatches: 134
Query Match: 31.33% Indels: 42
DB: 4 Gaps: 10

US-09-993-966-7 (1-470) x US-09-506-066E-7 (1-1285)
Qy      80 GluAspAspPheArgLeuGluValAlaLeuProProGluIuYrThrAspGly----- 96
Db      104 GACGACAAAGGTTCCCTGAGAAAGTTGCTGCTCCCTGAGAAAGTCTGAGGCGCATGAGGC 163
Qy      97 -----LeuGlySerGlyAspGluLysLysMetGluArgValSerGluProCys 112
Db      164 CAGGCGCAGCTTTCAGACAGATGATGGGAGAGGACCAAGCCGTGAGGTCCACTG 223
Qy      113 ProGlySerLysGlnLeuLysPheGluGluLeuGlnLysAspValSerMetGluGlu 132
Db      224 AGACTTACCAAGAGCACTTGAACTTGAAGCACTGACAGTGAAGTGTCTGATGAGAAAGAA 283
Qy      133 AspSerArgGlnGluTrpThrPheThrLeuTrpAspPheAspAsnAnglyLysValThr 152
Db      284 GACAAACCGCAAGATGAGATTCATTCATTAAGCTTGAACAACAGTGGAAAGTACAC 343
Qy      153 ArgGluAspIleThrSerLeuLeuHisThrIleTrpGluValAlaAspSerSerValAsn 172
Db      344 AGAGAGGACATGTCCAGCTGATGACACACATCTACAGAGGTGTGATGCTGTCAAT 403
Qy      173 HisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySer 192
Db      404 CACTCTCCGGCAGACAGCAACCTTCGAGTAAAGTAACTGACCTGAAACCTTCC 463
Qy      193 GlnSerLysArgSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArgProArgAla 212
Db      464 AGCAAGAGAGAAATGCTCTCTCACTGGCAAA---GACCGGAGAGCCCACTGCTGCAAGAA 520

```

```

Qy 213 GluThrIysProThrgluAspLeuArgSerTrpGluIuYsIsglnArgAlaProLeuArg 232
Db 521 GAGATTGAGCTCACAGATGAGAGCCCGGAGTGGCTGACGAGAGATATCCGCTTACAGCAGG 580
Qy 233 PheGlnIYAspSerArgLeuGlnIserGlyCySerYrHsIscYsValAspGluAsn 252
Db 581 AAGCCCAATGCTGATCCCGACGCTGCTGTGTGGAGAGCCCTATCTGTGTGATGAGAGAC 640
Qy 253 IleGluArgArgAsnHsIeYrLeuAspLeuAlaGlyIleGluAsnYrHsSerGlnPhe 272
Db 641 ACAGAGCGCAGAAACCACTACCTGTGACCTGTGTGATCGAGAACTACACATCTCAAGTTT 700
Qy 273 GlyProGlySerProSerValAlaGlnIuYsSerGluLeuProProArgHsSerAsn--- 291
Db 701 GGTCTGAGTGCACCACTGACGACGCGCAGGCAACATCATGCGAGGCGCACACACTT 760
Qy 292 ProThrgSerArgSerHsIsgluProGluAlaIleHsIleProHsIArgGlyProGln 311
Db 761 CCAAGCAGGTCCCGATCACAAAGATCGGATGCCACGCTATACACACCGCAGGTCCTCAA 820
Qy 312 -----GlyValAspProAlaSerPheHsIspHeuAspThrProIleAlaIys 327
Db 821 GTCTGGCTGAGCATGTCTATACAGCT-----AATGAGCCTGCCACCGCG 865
Qy 328 ValSerGluLeuGlnIArgLeuArgGlyThrGlnAspGlySerIYsHsPheValArg 347
Db 866 GCCCTGGCTGCACGCCCCGGATCAAGGGGCGAGAG-----AAGCACTTCTCAGG 916
Qy 348 SerProIYsAlaGlnIYsSerValGlyValGlyHsValAlaArgGlyAlaArgAsn 367
Db 917 TCTCTAAGGGGTCCAGGAAACCTCTTGGACA-----CCAGGCGAGTGGC 961
Qy 368 LysProProLeuGlyProAla-----IleProAlaVal 378
Db 962 AAGCCA-----GGAAAGCTCTCAGCTATTGCTGACGCGCTGCCATTCGCCAGAGT 1015
Qy 379 SerProSerAlaHsIleuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGly 398
Db 1016 GCTCAGAGATGGCCACCACTTCTCAGACCCCAACACAGCTTCCACCGCAGCCCTTATGCT 1075
Qy 399 HisIleYsHsIeYrHsIArgAlaIYsGluSerGlnIsglnIYsArgGlyLeuGlnAla 418
Db 1076 CACAAAGGGGTACCGGAGAAAGCCAGAGAGCCACATCACCACTTAAGGGCGATGGCCAG 1135
Qy 419 ProLeuAlaSerGlyGlyProValLeuGlyArgGluHsIleuArgGlyLeuProAlaLeu 438
Db 1136 CCT-----ACCATGATGGAGCATGAAGTACGCGACCTGCCCTCCCATG 1180
Qy 439 ValValYrGluSerGlnAlaGlyGlnProValGlnArgHsIsgluHsIeHsIeHs 458
Db 1181 CTGGGGCGCTGAGGGCTATGTATGCTGTGTGTCAGAGGCTATGAACACCACTATCACAT 1240
Qy 459 GluHisIeHsIeYrHsIeHsIeHsIeHsIeHsIeHsIeHsIeHsIeHsIeHsIeHs 466
Db 1241 GAGCACCACCACTATCACCAAC 1264

RESULT 4
US-09-506-066E-9
; Sequence 9, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9

```

```

; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-506-066E-9

Alignment Scores:
Pred. No.: 4,95e-41 Length: 1307
Score: 583.50 Matches: 144
Percent Similarity: 54.25% Conservative: 41
Best Local Similarity: 42.23% Mismatches: 103
Query Match: 23.35% Indels: 53
DB: Gaps: 10

US-09-993-966-7 (1-470) x US-09-506-066E-9 (1-1307)

Qy 150 LysValThrArgGluAspIleThrSerLeuLeuHsIleThrIleYrGluValAlaAspSer 169
Db 52 GAGGTCAACCACTCTCGGCGACGACGACACCTCCGCTGTGAAGCTTACCGTCAAGCCT 111
Qy 170 SerValAsnHsIserProThrgSerArgSerIYsMetLeuArgValIYsLeuThrValAlaPro 189
Db 112 TCGGTCAACCACTCTCGGCGACGACGACACCTCCGCTGTGAAGCTTACCGTCAAGCCT 171
Qy 190 AspGlySerGlnSerIYsArgSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArg 209
Db 172 GAGCGCTCCAGCAAGAGAGAGAGGCTCTGTGTCGACGAGACCGGAGCCACCGCT 231
Qy 210 ProArgAlaGluThrIYsProThrgIuAspLeuArgSerTrpGluIuYsIsglnArgAla 229
Db 232 TGCAGATGAGGGGTGAACCTGGCAGAGAGCCAAAGGTGGCTACAGAGAGTGTCTGCA 291
Qy 230 ProLeuArgPheGlnIYsAspSerArgLeuGlnIserGlyCySerYrHsIscYsVal 249
Db 292 CAGGTCAAGAGGCCCACTGACACCCCAAGCCCTGCTCGAGGGGGGCTTACTGTG 351
Qy 250 AspGluAsnIleGluArgArgAsnHsIeYrLeuAspLeuAlaGlyIleGluAsnYrHs 269
Db 352 GACGAGAAACGAGGAGCGCAGAAACCACTACCTGACCTCCCGGATTTGAACTACAG 411
Qy 270 SerGlnPheGlyProGlySerProSerValAlaGlnIYsSerGluLeuProProArgThr 289
Db 412 TCCAGATTCGGCCCTGTGCTCTCTCTGTCAGCAAGAGAGAGAGCCCGCAGGCGAGGCC 471
Qy 290 SerAsn---ProThrgSerArgSerHsIeGluProGluAlaIleHsIleProHsIArg 308
Db 472 TCGCACTCCAGGCGCGGTCCGCTCCAGAGACCAATACATGCCGTACACACACCGC 531
Qy 309 LysProGln-----GlyValAspProAlaSerPheHsIspHeuAspThrPro 324
Db 532 AGGTCAAGGTGCTGTGGAACAGTGTGTCAGCTCG-----GAGCCT 576
Qy 325 IleAlaYsValSerGluLeuGlnIArgLeuArgGlyThrGlnAspGlySerIYsHs 344
Db 577 GCTGCCCGGCGCTGACAGACGCGCGCCGAGAGGGCGGAG-----AAGCAG 627
Qy 345 PheValArgSerProYsAlaGlnIYsSerValGlyValGlyHsIeValAlaArgGly 364
Db 628 TTCTCAAGTCCCCCAAGGGCTCCGGAAAGCC----- 660
Qy 365 AlaArgAsnIYsProProLeuGlyProAlaIleProAlaValAspProSer-----Ala 382
Db 661 -----CTGGGGTCCAGCCAGCAGCAAGTCCGGGAAAGCC 696
Qy 383 HisLeuAlaAlaSerProAlaLeuLeuProSerLeuAla----- 395
Db 697 TTCACTACTACTCTGCGGCGCTGCGCCCGCAGAGCCCTTACAGAGCGCCACCACTC 756
Qy 396 -----ProLeuGlyHsIeYsHsIeYsHsIeYsHsIeYsHsIeYsHsIeYsHsIeYsHs 409
Db 757 CCGCAGCCCCCAGCCACCTTACGGCCACAAAGCGTACCGCCAAAGGGCAGGAG--- 813
Qy 410 GlnGlnIYsYsArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValIleuGlyArg 429

```

```

Db          814 -----GGGCACCTGCCACTCAAGGCCCCCACAAGCTCAGCGCTGCACAGT-----      858
Qy          430 GLuhis-----LeuaYgluLueProAlaeValvalTYrGluseGrMalaglycin    447
            |||||
            :::::|||||
Db          859 GAGCACGAGGTGGTGCCGGGACCCTGCCGCCACGCGACGAGAGAGGGCTACGCGGTGCCA    918
            |||||
Qy          448 ProValGLnaRGHtsgLUhIeSHtshIshIsghluhtshIshIstYrChIshIsphE     467
            :|||
            ::|||
Db          919 GTATTCcAcGcgSgCaGaCcAccAcCaCcAcCaCcAcCaCcAcCaCcAcCaCcAcCaC   978
            :|||
Qy          468 TYr 468
            ::
Db          979 CAC 981

RESULT 5
US-09-506-066E-1
; Sequence 1, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4954
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (966)...(3752)
; US-09-506-066E-1
```

Alignment Scores:		
Pred. NO.:	2,15e-08	Length: 4954
Score:	215.50	Matches: 129
Percent Similarity:	33.00%	Conservative: 69
Best Local Similarity:	21.50%	Mismatches: 187
Query Match:	8.62%	Indels: 216
DB:	4	Gaps: 26

US-09-993-966-7 (1-470)	x	US-09-506-066E-1 (1-4954)
QY	41	GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle--- 59
DB	1281	CAGGATGNGAGCAGCGGGCGAGCCATGACGAGATCGGCATCAGCAGCATCTTCCAA 1340
QY	60	-----GlyArgSerThrArgGlnLeuValGlyAspValLeuArgAspThrLeuSer 76
	1341	GGCAGACGCGCAAAATACTCAAT-----ATGCAGCAGCAACTCGCG 1382
QY	77	GlnGlnGlnGlnAlaAspPheArgLeuGlnValAlaLeuProGluLysThrAspGly 96
DB	1383	CAGGATGAGAT-----GTGGTGATCCGCGCTGCCAGATGCGACGACAGCAGAC 1433
QY	97	LeuGlySerGlyAspGlnLysLysMetGlnArgValSerGluProCysProGlySerLys 116
DB	1434	ACTGGCCACGCGCCACTCGCGCACTGCACCTGCACCAACCAAGAGG----- 1475
QY	117	LysGlnLeuLysPheGlnGlnLeuGlnCysAspValSerMetGlu-----GlnAsp 133
DB	1476	GAGCGCATCCGACCTGAGAGAAATTACCTCGCACTGTCTCGTGGAGGGGGCGCAAGTCATCG 1535
QY	134	SerArgGlnGlnIuTrpThrPheThrLeuLysAspPheAsp---AaaAsnGlyLysValThr 152
DB	1536	CAGCGCGCTCAGGTTCTCGTTCACCTTTCACACCTGCGAGCGGCACTACGCGCAAAATTACA 1595

QY	153	ArgGlnuPrIleThrSerLeuHisThrLysGluValValAspSerValAsn	172
DB	1596	AAGGACGACATCGTGGGCGATGTGTACACCATATACAGATTCATTGGCAAGTGGGTG	1655
QY	173	HisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySer	192
DB	1656	GTGCCCACTGCGGCGACGAAAGACAAATACAGTGGGCTCACCGTCAATGCCAGGGGCAA	1715
QY	193	GlnSerLysArg-----	196
DB	1716	TCGAAATATGACGCGGATGTGTCGCGTTCCGGTGGCAGCCGGATTACAGACGACCGCC	1775
QY	196	-----	196
DB	1776	AGCAAACTGAAGAGTTGCCACGAGGTGTGGCGGCATTCTGAAACCCCTGGCGCGGA	1835
QY	197	-----SerValLeuValAsnGlnAlaSerLeuGlnSerAla---	208
DB	1836	GGAATGGGATCCGGCGGAGCGTCCGGCTAACACATCCGCGGCAACCGCCGACAT	1895
QY	209	-----ArgProArgAlaGluThrLys---ProThrGluSerLeuArgSerTyrGluLys	225
DB	1896	CGTATGACACACGCAAACTGATTAAAGTCGATGACAGAGACGATGACAGCAACAGCGAA	1955
QY	226	LysGlnLysArg-----AlaProLeuArgPheGln-----	234
DB	1956	AAGGAGAGGACGCGCGCCACGCGCTCCGCGCGACGACCCGACCGAGATGGAACAAAG	2015
QY	235	-----GlyAspSer-----ArgLeuGlnGlnSerGlyCysTyrHisHis-	247
DB	2016	GCGACTGGGAAGACCATCACACCACTGCGATGCCGACGTATTCACACGAAGAACAT	2075
QY	248	-----CysValAspGluAsnIleGluArgArgAsnHis-----Tyr	259
DB	2076	TCCGGGGGCGAGCAGTGTGTGACGGAACAGAAATACGCCCGACATGGCACAATATAC	2135
QY	260	LeuAspLeuAlaGlyIleGluAsnTyrThrSerLysPheGlyProGlySerProSerVal	279
DB	2136	GAGAAATATGCGAATCTCAAGTGTGCGAAGCCGAGGAGT-GGACCGAGTGGACTGTCTCTC	2194
QY	280	AlaGlnLysSer---GluLeuProProArgThrSerAsnProThrArgSerArgSerHis	298
DB	2195	GCACCGACAGCACACACAGAGCCACCGAACCATCAATGCGCGCACAGACATCTAAT	2254
QY	299	GluPro-----GluAlaIleHisIlePro---	306
DB	2255	GAAACAGGCCACCCAGCGGGTCAAGATGTTGCCAAGGGCGGCAACAAAGTACACAGA	2314
QY	307	-----HisArgLysProGln	311
DB	2315	CCACTGCCTCGAAACGCGACGACGCGACGCTGTCAGTGGGCAACGATTCGCGTCCGGA	2374
QY	312	GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeu	331
DB	2375	TGCGGCATCTGCGAGTGCAGACGACGCGCGCGGTGGTACACCCGACCGCAAGTGTGAACCA	2434
QY	332	GlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysAla	351
DB	2435	CAGAGGCC-----	2443
QY	352	GlnGlyLysSerValGlyValGlyHisValAlaArgGlyAla-----ArgAsnLysPro	369
DB	2444	-----GTCGGGGGTACACACCGCTGGGCGGTGGGGGTGGTGGCGACATGATCT	2491
QY	370	ProLeuGlyProAlaIleProAlaValAspProSerAlaHisLeuAlaIleAspProAla	389
DB	2492	CGATGGGATGCGAGTGGCGTACGCGGACCCCATTCCTCACCCGACGACATCAACA	2551
QY	390	LeuLeuProSerLeuAlaProLeuGlyHisLys-----Lys	401
DB	2552	GCAAAATCAGACGACGACGACGACGACGCAAAATCGGCGGAGTGGGAAATCGCGCT	2611
QY	402	HisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu---	420

```
Db 2612 GAATCGCAAGATTATTAATGATCATCAGGAGAGATGAGAGAACCGCTGTGTTT 2671
Qy 421 -----AlaSerGlyGly----- 424
Db 2672 TCAGCTGAATGAGAAACCCAGCCAAATGTGATGCCATTCAGCAACCGGAGAGCAACA 2731
Qy 425 -----ProValLeu-----GlyArgGlyHisGlyLeu 435
Db 2732 ACAACCAACAACAGCAGACGCTGCATATACGGCTCGAATAATCCAC----- 2785
Qy 436 ProAlaLeuValValTyrGlySerGlnAlaGlyGlnProVal----- 449
Db 2786 -----GTTAATTATCAACACCAAGTCGGTCCGACAGAGTCGCGCTCAGCTCAG 2836
Qy 449 ----- 449
Db 2837 TCACACCCAGCGAGGCCCAACCCCAAGCATTCACAGACTCCGCGGCATCGAGT 2896
Qy 450 GlnArg-HisGlnHisHisHisGlnHis-----HisIleTyrHisHis 466
Db 2897 CAACGGTCAGACGACCAACCCCACTCATCCACTATCCAGCAGCAACCAACACAC 2954

RESULT 6
US-09-270-767-14537
; Sequence 14537, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1998-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14537
; LENGTH: 2063
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14537

Alignment Scores:
Pred. No.: 0.000322 Length: 2063
Score: 160.50 Matches: 81
Percent Similarity: 37.62% Conservative: 36
Best Local Similarity: 26.05% Mismatches: 97
Query Match: 6.42% Indels: 99
Gaps: 14

US-09-993-966-7 (1-470) x US-09-270-767-14537 (1-2063)
Qy 122 GlnGluLeuGlnCyAspValSerMetGlu-----GluAspSerArgGlnGluTyr 138
Db 1142 CAGGAATTACACCTGCGACGTGTCGTGAGGGCGCAAGTCATCGACCGCTGCATTC 1201
Qy 139 ThrPheThrLeuTyrAspPheAsp---AsnAsnGlyValValTrrArgGluAspIleThr 157
Db 1202 TCGTTCACGTTTACGACCTGACGCGGCAATCAACGGAATAAAGAGACATCGTG 1261
Qy 158 SerLeuHisThrIleTyrGluValValAspSerValAsnHisSerProThrSer 177
Db 1262 GCGTTTGTCACACCATATACGATCGATTGGCAAGTCGCGTGGTGGCCCATCGCGCC 1331
Qy 178 SerIleMetLeuArgValIleLeuThrValAlaProAspGlySerGlnSerIleArgSer 197
Db 1322 AGCAAGACATCAACAGCGCGCTCACCGTCACTCCGAGGGCAATCGAATCGACGCG 1381
Qy 198 ValLeu----- 199
Db 1382 GTGGTCCCGCTTCGGTGGCAGCGGATTCAGCAGACCAACCGCAGCAACTGAGAGAG 1441
Qy 199 ----- 199
```

```
Db 1442 TTGCCACGGGTCTGGCGCCATGTGCMAACGCTGGCCGCGAGAGTGGATCCGGC 1501
Qy 200 -----ValAsnGlnAlaAspLeuGlnSerIleArgProArgAla 212
Db 1502 GAGCTGCGGCTAACGACATCCGCGCAACCGCGCCAGCATGCTATCGACACGCAAA 1561
Qy 213 GluThrIle---ProThrGluAspLeuArgSerTyrGluIleValGlyGlnArg----- 228
Db 1562 CTGATTAAATGCTGATGACGAGACGATGACAGCAACGCAAGAAAGAAAGAACGCGCC 1621
Qy 229 ---AlaProLeuArgPheGln-----GlyAspSer--- 237
Db 1622 CAGCGCCCTCCCGCCGACCGCCAGCGGAAGTGCMAAAGGCGACTGGGAAGACCAT 1681
Qy 238 ---ArgLeuGlnGlnSerGlyCysTyrHisHis----- 247
Db 1682 CACCACCAAGTCGAGTCCGCGAGGTATCAACAAAGAAACAATTCCGGGCGAGACATGC 1741
Qy 248 CysValAspGluAsnIleGluArgArgAsnHis-----TyrLeuAspLeuAlaGlyIle 265
Db 1742 TCACGGAACAAGAAATACGCCCCGACATGGCCACAAATACGAGATATGCTAATCTC 1801
Qy 266 GluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla----- 280
Db 1802 AAGTGTGCAAGCCGAGGT-GGACCAAGTGTGACTGTCCCTCGCACCCAGACACCA 1860
Qy 281 -----GlnIleSerGluLeuProProArgThrSerAsn-----ProThr 293
Db 1861 GAGCCACCCCAACCATCAAAATGCGCAGC-AGGACATCTACATGAGCAGGCCACCAAC 1919
Qy 294 ArgSerArgSerHisGluProGluAlaIleHisIleProHisArgIleProGln-----Gly 312
Db 1920 GCGTCAAAGATGTTGCAAGGCGCGCA-----AACAAAGAAACCCCAAGCCAT 1970
Qy 313 ValAspProAlaSerPheHisPheLeuAspThrProIleAlaIleValSerGluLeuGln 332
Db 1971 GTGAGTCCC-----ATACGGCAACCGCGCAGACAAACAACCAACCAACAA 2015

RESULT 7
US-09-252-991A-15134
; Sequence 15134, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15134
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15134

Alignment Scores:
Pred. No.: 0.00084 Length: 1449
Score: 153.00 Matches: 118
Percent Similarity: 32.89% Conservative: 55
Best Local Similarity: 22.43% Mismatches: 189
Query Match: 6.12% Indels: 164
Gaps: 28

US-09-993-966-7 (1-470) x US-09-252-991A-15134 (1-1449)
```

```

QY 14 ArgArgGluserProgluGlyAserPheAlaValSerAlaAlaTrrAlaArgGly 33
DB 16 CGTGGCGCAAGACGGGAGGTGAGTCAAGAAACAACTGAGCTGAGAAAGCGGAGT 75
QY 34 IlegluGluTrpIlegluArgGlnArg-----CysProGly 45
DB 76 GCTCGACAGACTTTCATGAGCTGGGCAAGCTGGCACTTTCGAAAGACCTTGCA 135
QY 46 GlyValSerGly----- 49
DB 136 GAGCTGAGAAACCGGCGCATGTGTCCTGACGTCAAGGCAACATGATGAGAT 195
QY 50 -----ProArgGluLeuAlaGly-----ThrIlegluArgSerThrArgGluLeu 66
DB 196 TTCCCAACCCCATGTCTTCGAGACATGCGGTGAGCGTCTACTACAAACGCTTCGCAAA 255
QY 67 ValGlyAserValLeuArgAerThrLeuSerGluGluGluAerPhe---ArgLeu 85
DB 256 GTGGGCGCAACTTTCGAAAGACTCGGCGTGAACCCGAAACCGGTATCTCCAGCGCTTA 315
QY 86 -----GluValAlaLeu-----ProProGluGlyThrAsp 95
DB 316 CGACAAAGATCAAGTCTGCGCGCTCCAGCAAGAAAGATCTCCACAGACATCCAGCA 375
QY 96 GlyLeu-----GlySerGlyAerGluValLeuGlyMetGluArgValSerGluPro 111
DB 376 GGGTGTACAGCAACCGCGGCAAAATGGCGATGTGACCTCCGTGAAAGCGCTCAACACT 435
QY 112 CysProGlySerValLeuGluLeuGlyPheGluGluLeuGlnCysAerValSerMetGlu 131
DB 436 GCAACATTCGAGGAGCTCATGTCGACGCGCTGATCCGAGTGAATCCGCAACTCCGG 495
QY 132 GluAerSerArgGlnGluTrpThrPheThrLeuGlyAerPheAerAerAerGlyVal 151
DB 496 CCAAGATGTGGGCGAGGACGCGCAAGCAAGAAACACCAAGCGGTATATGCGAAAGCAC 555
QY 152 ThrArgGluAerIleThrSerLeuLeuHisThrIleGluValValAerSerSerVal 171
DB 556 CTACGCGCG----- 564
QY 172 AeriHiserProThrSerSerLeuMetLeuArgValLeuLeuThrValAlaProAer-- 190
DB 565 ---CATCTACACAGATGATCATCTTCGAAACCAAGCGGCGCTTCGACCGGCAAC 621
QY 191 ---GlySerGlnSerLeuAerSerValLeuValAerGlnAlaAerGlnSerAlaArg 209
DB 622 CATGGGCAAGGTCGCGCACTCGGCGCT-----GATGGCGCGAGAA 660
QY 210 ProArgAlaGluThrLeuProThrGluAerLeuArgSerTrpGlu----- 224
DB 661 GCGCGAGAAATACGCGTCCGACAGAAAGACTTCGAAATGACGCGGACGCGACATGCG 720
QY 225 -----LeuValGlnAerAlaProLeuArgPheGlnGlyAer 236
DB 721 CGTGGCTGGGCGGAGCGGCGGTGATGACAGCAAGCTGAGAGCGGCGGCAATCTG 780
QY 237 SerArgLeuGlnLeuSerGlyCysValGlnHisCysValAerGluAerIlegluArg 256
DB 781 GCGCGCTGCGCAACGAGAA-----CGCAC 807
QY 257 AeriHisThrLeuAerLeuAlaGlyIlegluAerThrSer----- 270
DB 808 GATCCGCGAGTCTGAGTCAAGTGGCGCGTCAACCGGCGCATCCGATCCGCGGAGAT 867
QY 271 ---GlnPheGlyProGlySer-----ProSerValAlaGlnLeuAerGluLeuPro 286
DB 868 CTTCGCGCTGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 927
QY 287 ProArgThrSerAerProThrArgSerArgSerHisGluProGluAlaIleHisIlePro 306
DB 928 CCGCAAGAGCAAGCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 984

```

```

QY 307 HisArgGlyProGlnGlyVal---AerProAlaSerPheHisPheLeu----- 321
DB 985 CATTCGCGTACAGATGAGACGCTGATCCGCGCAAGACACCATCTTCGTCGACCGGCAA 1044
QY 322 -----AerThrProIleAlaLeu-----ValSerGluLeu---Gln 332
DB 1045 CGTCTCGCGGAGTACTCTGACCGACCTGTTCCGATCATGAGAACTGGGCACTCGGCGCAA 1104
QY 333 GlnArgLeuArgGlyThrGlnAerGlySerLeuHisPheValAerSerProGlyValAlaGln 352
DB 1105 GATGCTCTCATGAGTTCGCGTATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1164
QY 353 GlyLeuSerValGlyValGlnHisValAlaArg-----GlyAlaAerGlnLeuPro 369
DB 1165 -----CTCGGCAACCGCAAGCAAGTCCAGCAACCTGTGAGAAAGAACTACTCT 1209
QY 370 ProLeuGlyProAlaIleProAlaValAerProSerAlaHisPheValAerProAla 389
DB 1210 GCGCTGGGA-----CTCCCTGGGCGGAGTCTCTGGC----- 1239
QY 390 LeuLeuProSerLeuAlaProLeuGlyHis----- 399
DB 1240 -----CTTGCGCGTGTCTTGAGAGAAACCGGATCAAGACCGGCAACGCGCAAGCGCAA 1293
QY 400 -----LysLeuHisPheHisPheHisPheHisPheHisPheHisPheHisPheHis 410
DB 1294 GCTGCTCGCGAGCGCGCTGAGCAAGCGGCAAGCGGCAAGCTGTGAGCAACAAAGTCCGC 1353
QY 411 -----GlnGlyCysAerGlyLeuGln-----AlaProLeuAlaSerGly----- 423
DB 1354 GTGCGCGAAGTGGGAGCATGCAACACCGGCGAGCGCACTTCTACTGCGAGTACTG 1413
QY 424 GlyProValLeuGlyArg 429
DB 1414 GCGCCAGCGCGCTGCGCGC 1431

```

RESULT 8
 US-09-252-991A-14877
 / Sequence 14877, Application US/09252991A
 / Patent No. 6551795
 / GENERAL INFORMATION:
 / APPLICANT: Marc J. Rubenfield et al.
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 / FILE REFERENCE: 107196.136
 / CURRENT APPLICATION NUMBER: US/09/252,991A
 / PRIOR FILING DATE: 1999-02-18
 / PRIOR APPLICATION NUMBER: US 60/074,788
 / PRIOR FILING DATE: 1998-02-18
 / PRIOR APPLICATION NUMBER: US 60/094,190
 / NUMBER OF SEQ ID NOS: 33142
 / SEQ ID NO 14877
 / LENGTH: 2226
 / TYPE: DNA
 / ORGANISM: Pseudomonas aeruginosa
 / US-09-252-991A-14877

Alignment Scores:
 Pred. No.: 0 0016 Length: 2226
 Score: 153.00 Matches: 118
 Percent Similarity: 32.89% Conservative: 55
 Best Local Similarity: 22.43% Mismatches: 189
 Query Match: 6.12% Indels: 164
 DB: 4 Gaps: 28

US-09-993-966-7 (1-470) x US-09-252-991A-14877 (1-2226)
 QY 14 ArgArgGluserProgluGlyAserPheAlaValSerAlaAlaTrrAlaArgGly 33
 DB 603 CGTGGCGCAAGACGGGAGGTGAGTCAAGAAACAACTGAGCTGAGAAAGCGGAGT 662
 QY 34 IlegluGluTrpIlegluArgGlnArg-----CysProGly 45

```

Db      663 GCTCGACAGCATGTTCAATGAGCTGCGGCAAGCTGCGGACTTCTTCGAGAAGACCTTGCA 722
      46 G1ValSerGly-----
Db      723 GGAATCGAAGAAACCGCGCTCATGTGTCCTTCAGCTCAAGGCGACCATGATGAAGAT 782
      50 -----ProArgGlnLeuArgLeuAlaGly-----Thr1LeuGlySerThrArgGlnLeu 66
      783 TTCCCACTCCGATCGTTCGACATGCGGTGAGCGGTCTACTACAGGACGCTTTCGACAA 842
      67 ValG1ValSerValLeuArgAspThrLeuSerGlnGlnGlnLeuAspAspPhe---ArgLeu 85
      843 GTGGGGCAACTCTTCGAGAAGAGCTGCGGCGTGAACCCGAAACAAGGTTCTCCACGCTTA 902
      86 -----GluValAlaLeu-----ProProGlnLysThrAsp 95
      903 CGACAAAGATCAAGTCGTGCGCGCTCCAGCAAGAAAGATCCTCCACGACATCCACGA 962
      96 G1Leu-----GlySerG1ValProGlnLysMetGlnArgValSerGlnPro 111
      963 GGTGTACAGCCACCGCCCGAATAATGGCATGTGCTGCTGTAAGGACATCACCAACT 1022
      112 CysProGlySerLysGlnLeuLysPheGlnGlnLeuGlnCysAspValSerMetGln 131
      1023 GCACATTCGACGACGCTCATCGTCGACGCTCGATGCGCGCGATGATCCGCAACTCCG 1082
      132 GluAspSerArgGlnGlnLysThrPheThrLeuTyrrAspPheAspAsnAsnGlyLysVal 151
      1083 CCAGATGTGGGGCGAAGACGCGCAAGCAAGAACACCAAGCGGGTAATGCGGAAAGCAC 1142
      152 ThrArgGlnAspLysIleThrSerLeuHisThrLysIleTyrrGlnValValAspSerVal 171
      1143 CTACGCCG-----
      172 AsnHisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAsp--- 190
      1152 ---CATTACCAAGAGATGATCACTCTGCAACAACGAGCGGCTTCGACCGGACAC 1208
      191 ---GlySerGlnSerLysAspSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArg 209
      1209 CATGGGAGCGGTGCGCAAGCTCGGCT-----GATGGCGCAGAA 1247
      210 ProArgAlaGlnThrLysProThrGluAspLeuArgSerThrGlu----- 224
      1248 GGCAGGAATACGGCTCCACGACAAAGACTTGAATGACCGCCGACGCGACCATGCG 1307
      225 -----LysLysGlnArgAlaProLeuArgPheGlnGlyAsp----- 236
      1308 CGTGCTCTGCGCGACGCGAGCGTGTGATGACAGACGACGCTGAGACCGCGGACATCTG 1367
      237 SerArgLeuGlnGlnSerGlyCysTyrrHisCysValAspGlnAsnLysLeuArgArg 256
      1368 GCGCGCTGCGCAACCAAGG-----CGCACCC 1394
      257 AsnHisTyrrLeuAspLeuAlaGlyLysLeuAsnTyrrThrSer----- 270
      1395 GATCCGCACTGGGTCAAGTGGCGGTGACCCGCGCACGCGCATCCGATACCCGGGGAT 1454
      271 ---GlnPheGlyProGlySer-----ProSerValAlaGlnLysSerGlnLeuPro 286
      1455 CTTTGGCTGACCGCGGACGCGCGCCGCGACCGGAACTGGGCAAGAAAGTCGAGCTGA 1514
      287 ProArgThrSerAsnProThrArgSerArgSerHisGlnProGlnAlaIleHisIlePro 306
      1515 CCTCAAGGACCAACACGATGACCGGTCTGCA---CATCAGCATCATGGGCTTCAACGAGGC 1571
      307 HisArgLysProGlnGlyVal---AspProLysSerPheHisPheLeu----- 321
      1572 CATTCGGTCAGCATGGAACCTGATTCGCGGACGAGACACCATTCGTGGTACCGGCA 1631
      322 -----AspThrProIleAlaLys-----ValSerGlnLeu---Gln 332

```

```

Db      1632 CGTCTGCGGCACTACCTGACCGACCTGTTCCCGATCATGGAATGGGCACTCGGCCAA 1691
      333 GlnArgLeuArgGlyThrGlnAspArgLysLeuHisPheValArgSerProLysAlaGln 352
      1692 GATGCTCTTCATCGTTCCTGCTGATGGCGGCGCGGACGATGACGAACCGGCGCGGCGG 1751
      353 G1LysSerValGlyValGlnHisValAlaArg-----G1ValArgAsnLysPro 369
      1752 -----CTCGGCAACCAAGCACGCTCCAGCAACTGTGGAGAAAGAACTACTCT 1796
      370 ProLeuGlyProAlaIleProAlaValSerProSerAlaHisLeuAlaIleAspProAla 389
      1797 GCGCTGGGA-----CTCCCTGGGCGGAGTTCCTGCG----- 1826
      390 LeuLeuProSerLeuAlaProLeuGlnHis----- 399
      1827 -----CTTGGCGGTGTCTCTGAGGAAGAACCGGAGATCAAGACCGGCAACCGGCCAA 1880
      400 -----LysLysHisLysHisArgAlaLysGlnSerGln----- 410
      1881 GCTGCTCGGCAAGCGCTGACGACGACGCGGCAAGCTCTGGACAAACAAGTGGCC 1940
      411 -----GlnGlyCysArgGlyLeuGln-----AlaProLeuAlaSerGly----- 423
      1941 GTGCGCGAAGTGGCGGACATCGACAACCGGCGACGCACTTCACTGCGGATGTACTG 2000
      424 G1ProValLeuGlnArg 429
      2001 GGCCCAAGCCCTGCGCGC 2018

RESULT 9
US-09-252-991A-14487/C
; Sequence 14487, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14487
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14487

Alignment Scores:
Pred. No.: 0.00165 Length: 2268
Score: 153.00 Matches: 118
Percent Similarity: 32.89% Conservative: 55
Best Local Similarity: 22.43% Mismatches: 189
Query Match: 6.12% Indels: 164
DB: 4 Gaps: 28

US-09-993-966-7 (1-470) x US-09-252-991A-14487 (1-2268)
      14 ArgArgGlnSerProGlnGlyAspSerPheAlaValSerLysAlaIleArgGlyGly 33
      1657 CGTCGGAAGACGCGCAAGGTGCAAGTCAAGAAAGCACTGAGCTCGAAGAGCGCAAGT 1598
      34 IleGlnLysProGlnGlyVal---AspProLysSerPheHisPheLeu-----CysProGly 45
      1597 GCTCGACAGCATGTTCAATGAGCTGCGGCAAGCTGCGGACTTCTTCGAGAAGACCTTGCA 1538
      46 G1ValSerGly-----
      1537 GGAATCGAAGAAACCGCGCTCATGTGTCCTTCAGCTCAAGGCGACCATGATGAAGAT 1478

```

QY 50 -----ProArgGlnLeuArgLeuAlaGly---ThrIleGlyArgSerThrArgGlnLeu 66
 Db 1477 TTCACACCCCATGCTTTTCGACATCGCGTGAAGCGCTTACCAAGACGCTTCGACAA 1418
 QY ValGlyAspValLeuArgAspThrLeuSerGlnGluGluAspAspPhe---ArgLeu 85
 Db 1417 GTGGGGCCAACTTTCGAAGAGCTCGGCGTGAACCCGAACACGGTATTCACAGCGTCTA 1358
 QY 86 -----GluValAlaLeu-----ProProGlnLysThrAsp 95
 Db 1357 CGACAGATCAAGTCGCTGCCCGCTCCAGCAGAAAGATCTCCACGACATCCAGCA 1298
 QY 96 GlyLeu-----GlySerGlyAspGlnLysMetGluArgValSerGlnPro 111
 Db 1297 GGTGTACAGCCAGCCCGCGGAAATGGGATGTCGACTCCGTGAAGGATCAACCAACT 1238
 QY 112 CysProGlySerLysLysGlnLeuLysPheGlnGluGluGlnCysAspValSerMetGlu 131
 Db 1237 GCACATTCGAGGAGCATTCGTCCAGCGCTCGATCCGCGCATTCGCAACTCCGG 1178
 QY 132 GluAspSerArgGlnGlnTrpThrPheThrLeuTyzAspPheAspAsnGlyLysVal 151
 Db 1177 CCAAGATGTGGGAGAGGACGAGCAAGCAAGAAAGACCAAGCGGTATTCGCGAAAGCAC 1118
 QY 152 ThrArgGluAspIleThrSerLeuLeuHisThrIleTyrgluValValAspSerSerVal 171
 Db 1117 CTACGCGCGG----- 1109
 QY 172 AsnHisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAsp--- 190
 Db 1108 ---CATTTACACAGAGATGATCAATCTTCGACAGACCAAGCGGCTTCGACCGGACAC 1052
 QY 191 ---GlySerGlnSerLysArgSerValLeuValAsnGlnIleAspLeuGlnSerAlaArg 209
 Db 1051 CATGGGAGCGTCCGACGACTCGGCT-----GATGGCGGACGAA 1013
 QY 210 ProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGlu----- 224
 Db 1012 GCGCGAGAAATACGCGTCCGACGACCAAGACCTTCGAATATGACCGCGACGCGACATGCG 953
 QY 225 -----LysLysGlnArgAlaProLeuArgPheGlnGlnAsp--- 236
 Db 952 CGTGTCTCTGGCGGAGCGGACGCGTGTGATGACAGCAGCAGGTGAGACCGGCGACATCTG 893
 QY 237 SerArgLeuGlnGlnSerGlyCysTyzHisIleCysValAspGluAsnIleGluArgArg 256
 Db 892 GCGCGCTGCGACCAACCAAGG-----CGCAC 866
 QY 257 AsnHisTyrlAspLeuArgLeuAlaGlyIleGluAsnTyrlThrSer----- 270
 Db 865 GATCCCGGACTGGCTGCAAGCTGCGCTGACCCGCGGACCGCGCATGATCCCGCGGCGAT 806
 QY 271 ---GlnPheGlyProGlySer-----ProSerValAlaGlnLysSerGlnLeuPro 286
 Db 805 CTTCCTGGCTGAGACCGGAGCGCGCCCAAGACCGGAACTCGGAGAAAGGTGCACTGTA 746
 QY 287 ProArgThrSerAspProThrArgSerArgSerHisIleGluProGluAlaIleHisIlePro 306
 Db 745 CCGCAAGAACCAAGACCTGACCGGTCTCGA---CATAGACATCATGGGCTACCAAGAGGC 689
 QY 307 HisArgLysProGlnGlyVal---AspProAlaSerPheHisPheLeu----- 321
 Db 688 CATTCGCTGACGATGAAAGCGCTGATCCGCGGCAAGACACCATCTCGGTGACCGGCAA 629
 QY 322 -----AspThrProIleAlaLys-----ValSerGlnLeu---Gln 332
 Db 628 CGTCTCGCGGAGATTAACCTGACCGACCTGTTCGCGATGAGAACTGGGCACTCGCGCAA 569
 QY 333 GlnArgLeuArgGlyThrGlnAspGlySerLysIlePheValArgSerProLysAlaGln 352
 Db 568 GATGCTCTCATGCTTCCGCTGATGGCGGCGGCGGATTAAGAAACGCGGCGCGCG 509

QY 353 GlyLysSerValGlyValGlyHisValAlaArg-----GlyAlaArgAsnLysPro 369
 Db 508 -----CTCGGACACCAAGACGCTCCAGCAACTGCTGAGAAAGAACTACT 464
 QY 370 ProLeuGlyProAlaIleProAlaValSerProSerAlaHisIleValAlaSerProAla 389
 Db 463 GCGCTGGGA-----CTCCCTGGCGGAGTTCTCGC----- 434
 QY 390 LeuLeuProSerLeuAlaProLeuGlyHis----- 399
 Db 433 -----CTGAGCGGTGCTCGAGAAACCGGAGATCAAGACCGGCAACGCCAGGCCAA 380
 QY 400 -----LysLysHisLysHisIleArgAlaLysGlnSerGln----- 410
 Db 379 GCTGCTCGCAAGCGCTGACGAGAGCCAGCCGCGAGCTCTGACAAACAAGTGGCC 320
 QY 411 -----GlnGlyCysArgGlyLeuGln-----AlaProLeuAlaSerGly----- 423
 Db 319 GTGGCGGAGGTGGCGGACATCGACAAACCGGCGGACGACCTTCTTACTGCGATGTA 260
 QY 424 GlyProValLeuArgArg 429
 Db 259 GCGCCAGCGCTGCGCGC 242
 RESULT 10
 US-08-999-689A-2
 ; Sequence 2, Application US/08999689A
 ; Patent No. 6541615
 ; GENERAL INFORMATION:
 ; APPLICANT: ULARICH, AXEL
 ; APPLICANT: KHARITONENKOV, ALEXEI
 ; APPLICANT: CHEN, ZHENGTUN
 ; TITLE OF INVENTION: SIRP PROTEINS AND USES THEREOF
 ; FILE REFERENCE: 038602/0548
 ; CURRENT APPLICATION NUMBER: US/08/999,689A
 ; CURRENT FILING DATE: 1997-11-14
 ; PRIOR APPLICATION NUMBER: 60/030,964
 ; PRIOR FILING DATE: 1996-11-15
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2433
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-08-999-689A-2
 Alignment Scores:
 Pred. No.: 0.00224 Length: 2433
 Score: 152.00 Matches: 140
 Percent Similarity: 33.04% Conservative: 51
 Best Local Similarity: 24.22% Mismatches: 189
 Query Match: 6.08% Indels: 198
 DB: 4 Gaps: 31
 US-09-993-966-7 (1-470) x US-08-999-689A-2 (1-2433)
 QY 3 LysLeuHisSerLysProAlaAlaValCysLysArgArgGlnSerProGlnGlyAspSer 22
 Db 507 GACCTTCACTGCGAGAGTCCGACGCGCTTTCACCCAGACATCACTGAA----- 557
 QY 23 PheAlaValSerAlaAlaTrpAlaArgLysGlyIleGlnGluTrp----- 37
 Db 558 -----ATGTTCAAAAATGGGAATAGCTTCAGA 587
 QY 38 IleGlyArgGlnArgCysPro----- 44
 Db 588 CTTCAGACCAAGCTGAGACCCCTAGAGAGACCGTCTCTACAGATCCACAGCACAGC 647
 QY 45 GlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSerThrArg 64
 Db 648 CAAGGTGTGTGATCCCGGAGGAGACGTTCACTCAAGTATCTGGAGGTGGCCACGT 707
 QY 65 GlnLeuVal-GlyAspValLeuArgAspThrLeuSerGlnGluGlnLysAspPheAsp 84

```

Db      708  CACCTTGACAGGGGAGACCTCTGTGGACCTGCCAATCTGTGAG----- 753
Qy      84  gLeuGlValAlaLeuProProGluThrThrArgGlyLeuGlySerGlyAspGluValSly 104
Db      754  ----ACCATCCGAGTTCACCC-----AC 773
Qy      104  sMetGluArgValSerGluProCysPProGlySerGlyLeuGluLeuLysPheGluGluLeu 124
Db      774  CTTGGAGGTTACTCACACGCCCGGTGAGGGGACAGAACAG-----GTGAATGT 821
Qy      124  uGlnCysAspValSerMetGluGluLysSerArgGlnGluTrpThrPheThrLeuTrpAs 144
Db      822  CACCTGCGAGGTGAGAAATTCTACCCCGAGAGACTACAGCTACAGCTGGTTG----- 873
Qy      144  pPheAspAsnArgGlyValThrArgGluAspGluLeuSerLeuLeuIleThrLeuIleTy 164
Db      874  ----GAGAAATGGAAACGTGTCGCCGACAGAAACGGCCTCA-----ACCGTTAC 917
Qy      164  rGluValValAspSerSerValAsnHisSerPro-----ThrSerSerLy 179
Db      918  AGAGAACAGAGATGTACTCAACAATCTGATGAGCTGCTCTGTGTAATGATATCTGCCCA 977
Qy      179  sMetLeuArgValLysLeuThr-----ValAlaProAsp-GlySerGlnSerLysArgS 197
Db      978  CAGGATGATGTGAAGCTCACTGCGCAGGTGAGACATGACGGGACGACGCGGTACAGCA 1037
Qy      197  eValLeuValAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysPro- 216
Db      1038  AAGCATAGACCTGAAAGGTCTCAG-----CCACCCGAGAGACAGGGCTCAATACCGCGC 1094
Qy      217  ----ThrGluAspLeuArgSer----- 222
Db      1095  TGAGAACACTGGATCTAATGAACGAAACATTAATTGTGTGTGGTGTGGTGTGACCTT 1154
Qy      223  ----TTP-----GluLysLysGlnA 228
Db      1155  GCTGTGGCCCTACTGATGGGGGCTCTACTCTGCTCGAATCAGACAGAAAGAACCCCA 1214
Qy      228  rGluArgProLeuArgPheGlnGly-----AspSerArgLeu-GluGlnSerGlyCys 244
Db      1215  GGGGTCCACTTCTTCTAACAAGTTGATGATGAGCCCGAGAAATGTCACAGAAATACACA 1274
Qy      245  ----TyrHisHisCysValAspGluValAsnGluLeuArgGluArgAsnHisLeu 260
Db      1275  GGACACAAATGATATCAATATGACAGACTGAACCTGCCCA----- 1316
Qy      261  AspLeuAlaGlyLeuAsnTyrThrSerGlnPheGlyProGlySerPro--SerVal 279
Db      1317  ----GGGGAAGAAAGCTGCTCCAGGCTGCGAGCCCAACAAACACACAGAGTA 1367
Qy      280  AlaglnLysSerGluLeuPro-----ProGlnThrSerAsnPro 292
Db      1368  TGGCGATTCACAGCCAGCCGCGAGCCGCTCGAGAGACACCTCACTAATGTGACCT 1427
Qy      293  ThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgGlyProGlnGly 312
Db      1428  ----GACATGTGCCA-----CCTCAACCGGAGACCCCAAGCA 1460
Qy      313  ValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGln 332
Db      1461  GCCGCCCCCAAGCC-----TGAGCCGTCTTCTCAGAGTACGC 1499
Qy      333  GluArg-----LeuArgGlyThr--GlnAspGlySerLysHisPhe-Val 346
Db      1500  CAGGTCACAGGTCCGAGAGAAATGAATGGAGCCGTGTGTTGCTGTAGACCACTCTTAC 1559
Qy      346  LArgSerProLysAlaGlnGlyLysSerVal-----Gly 357
Db      1560  GCGCTTTCTTGTCCACAGGAGCCCGCTGATATAGACACAGCCAAACCAAGTTCGCGAGG 1619
Qy      357  yValGlyHisVal-----AlaArgGlyAlaArgAsnLysProProLe 371

```

```

Db      1620  GCTGGGGCGGTGACAGCTCTGGAGCCAGGGGCGCTTCTCTCCCAACC-- 1677
Qy      371  uGluProAlaIleProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeu-L 391
Db      1678  ----CTCTTGGCTCTCCACACTTCC 1700
Qy      391  eProSerLeuAlaProLeuGlyHisLysHisLysHis-----ArgA 406
Db      1701  TGGGACGCCACGGCCCCCTC-----CCCCAATTGGCCACACCTGTGAGGTGACGTT 1754
Qy      406  LAlysGluSerGln-----GlnGlyCysArgGlyLeuGlnAlaP 419
Db      1755  GCCAAACACACCGGAGAACCAACCTGGAGAAATGGCCGAATCTGCGGGTCCAGAACT 1814
Qy      419  rLeuAlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGlu-LeuProAlaLeu 438
Db      1815  CTGTGCTCTCCGTCATCACTATGTGGTTTGAAGACCTCGACAGCTCCCGAGCT 1874
Qy      439  ValValTrpGlu-----SerGlnAlaGluProValGlnArgHis 452
Db      1875  CCGAAGCTGATCTTCCAGGGGTGGGAGAGAAATGCCACTCCCTGACTCCACAC 1934
Qy      453  GluHisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHisHisHis 466
Db      1935  CTCACACACACACACACACACACACACACACACACACACACACACACACAC 1976

RESULT 11
US-09-944-807-3
; Sequence 3, Application US/09944807
; Patent No. 6773895
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim Pharma KG
; TITLE OF INVENTION: Method for identifying substances which positively
; TITLE OF INVENTION: influence inflammatory conditions of chronic
; TITLE OF INVENTION: inflammatory airway diseases
; FILE REFERENCE: 082.00n
; CURRENT APPLICATION NUMBER: US/09/944,807
; PRIOR FILING DATE: 2001-08-31
; PRIOR FILING DATE: UK 0021484.1
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-944-807-3

Alignment Scores:
Pred. No.: 0.00224      Length: 2433
Score: 152.00      Matches: 140
Percent Similarity: 33.04%      Conservative: 51
Best Local Similarity: 24.22%      Mismatches: 189
Query Match: 6.08%      Indels: 198
DB: 4      Gaps: 31

US-09-993-966-7 (1-470) x US-09-944-807-3 (1-2433)
Qy      3  LysLeuHisSerLysProAlaAlaValCysLysArgArgGluSerProGluGlyAspSer 22
Db      507  GAGCTTCACTGGAGAGTCCACCGGCTTCTCACCGAGAGATCACTCCGAAA----- 557
Qy      23  PheAlaValSerAlaAlaIleTrpAlaArgLysGlyIleGluGluTrp 37
Db      558  -----ATGTTCAAAAATGGAAATGAGCTTCAGA 587
Qy      38  IleGlyArgGlnArgCysPro----- 44
Db      588  CTTTCAACACACAGTGAAGCCCGTAGAGAGAGAGGTCTCTACAGATCCACAGACAGC 647
Qy      45  GlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSerThrArg 64
Db      648  CAAAGTGTGTGACCCCGGAGAGACGTTTCACTCAAGTCAATGCAATGCGAGGTGCCACGT 707

```



```

QY      65 GlnLeuVal-GlyAspValLeuArgPheThrLeuSerGluGluGluLeuAspPheAr 84
DB      708 CACCTTGACGAGGAGACCTCTTCTGCGGACCTGCCAATCTTGCTAG----- 753
QY      84 GlnGluValAlaLeuProGluIuLyThrAspGlyLeuGlySerGlyAspGlyVal 104
DB      754 -----ACCATCCGAGTTCACCC-----AC 773
QY      104 smetGluArgValSerGluProCysProGlySerLeuGlnLeuLyPheGluGluLe 124
DB      774 CTGGAGAGTATCAACAGACCCCGAGGAGGAGAGAACAG-----GTAATGT 821
QY      124 uGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPheThrLeuTy 144
DB      822 CACCTCCGACGAGGAGGAGTTCACCCCGACGACTACGAGTGCCTGGTTG----- 873
QY      144 PheAspAsnAsnGlyValValThrArgGluAspIleThrSerLeuLeuHisThrIle 164
DB      874 -----GAGATGGAAAGCTGTCCTCCGACAGAAACGGCTCA-----ACCGTTAC 917
QY      164 rGluValValAspSerSerValAsnHisSerPro-----ThrSerSerly 179
DB      918 AGAGAACAGAGATGCTACCTACCAACCTGATGAGCTGCTCTGATGATGATGCTGCA 977
QY      179 smetLeuArgValIleLeuThr-----ValAlaProAsp-GlySerGlnSerlyAs 197
DB      978 CAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1037
QY      197 eValLeuValAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleAs 216
DB      1038 AACCGATGACCTGAAGCTCAG-----CCACCCGAAAGAGAGGAGGCTCAATACCGCGC 1094
QY      217 -----ThrGluAspLeuArgSer----- 222
DB      1095 TGAGAACACTGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1154
QY      223 -----Trp-----GluLyLeuGln 228
DB      1155 GCTGTGGCCCTTACTGATGCGCGCCCTTACTGCTCGATGATGATGATGATGATGAT 1214
QY      228 rGlnAlaProLeuArgPheGlnGly-----AspSerArgLeu-GluGlnSerGlyCys 244
DB      1215 GGGCTCCTACTCTTCTTACAAAGGTTGATGATGATGATGATGATGATGATGATGAT 1274
QY      245 -----TyrlHisHisCysValAspGluAsnIleGluArgAsnHisIle 260
DB      1275 GGCACAAATGATATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 1316
QY      261 AspLeuAlaGlyIleGluAsnTyrlThrSerGlnPheGlyProGlySerPro--SerVal 279
DB      1317 -----GGGAGAGAGAGCTGCTCCCGAGCTGCGAGCCCAACCAACGAGGAT 1367
QY      280 AlaGlnLySerGlnLeuPro-----ProArgThrSerAsnPro 292
DB      1368 TGCCACATTCACACAGCCCGGAGCCGCTGAGAGACACCTTCACCTATGCTGACT 1427
QY      293 ThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLyProGlnGly 312
DB      1428 -----GACATGCTCA-----CCTCAACCGGAGCCCGCAAGCA 1460
QY      313 ValAspProAlaSerPheHisPheLeuAspThrProIleAlaIleValSerGluLeuGln 332
DB      1461 GCGCGGCCCCAACCC-----TGACCGCTGCTTCTTCAGAGTACGCG 1499
QY      333 GlnArg-----LeuArgGlyThr--GlnAspGlySerLyHisPhe--Va 346
DB      1500 CACGCTCCAGGTCCTCCAGAGAGATGATGATGATGATGATGATGATGATGATGAT 1559
QY      346 IArgSerProLyAlaGlnGlyLySerVal-----Gly 357
DB      1560 GGCCTTTCTTGTCCACAGGAGCGCGCTGATGATGATGATGATGATGATGATGATGAT 1619

```

```

QY      357 ValGlyHisVal-----AlaArgGlyAlaArgAsnLyProProLe 371
DB      1620 GCTGGGCGGCTGACAGGCTTGGAGACCCAGGAGCCAGGAGGCTCTTCTTCCCAACC- 1677
QY      371 uGlyProAlaIleProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeu 391
DB      1678 -----CTCTTGAGCTCTCCAGCATTTCC 1700
QY      391 euProSerLeuAlaProLeuGlyHisLyValHisLyHis-----ArgA 406
DB      1701 TGGGACAGCACAGGCGCCCTC-----CCCAACATTTCCACACACTGAGAGCTGACT 1754
QY      406 IalysGluSerGln-----GlnGlyCysArgGlyLeuGlnAlaP 419
DB      1755 GCCAACACCCAGGAGGAGAACCAACCTGGAGAGTGGCCAGAACCTGCTGGGCTCAAGAACT 1814
QY      419 rGluAlaSerGlyGlyProValLeuGlyValArgGlnHisLeuArgGlu--LeuProAlaLeu 438
DB      1815 CTGTGCTCTCCGTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1874
QY      439 ValValTyrlu-----SerGlnAlaGlyGlnProValGlnArgHis 452
DB      1875 CCGAAGCTGATCTTCCAGGAGGAGGAGAGAGAAATCCACCTCCCTGACCTCACAC 1934
QY      453 GlnHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 466
DB      1935 CTCACACACACACACACACACACACACACACACACACACACACACACACACAC 1976

RESULT 12
US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 296 Length: 4403765
Score: 149.50 Matches: 133
Percent Similarity: 29.76% Conservative: 67
Best Local Similarity: 19.79% Mismatches: 200
Query Match: 5.98% Indels: 272
DB: 3 Gaps: 30

US-09-993-966-7 (1-470) x US-09-103-840A-2 (1-4403765)
QY      9 AlaAlaValCysValArgArgGluSerProGluGlyAspSerPheAlaValSerAlaAla 28
DB      3034730 GCCGCCATACCCGACGCGCTCGTGCACCA----- 3034759
QY      29 TrpAlaArgLyGlyIleGluGluTrpIleGlyArgGlnArg-----CysProGlyGly 46
DB      3034760 ---GTCGTGCGCGCTCCGAGAGCTGAGCACACCGGCGTGCAGACCGGAGTGT 3034816
QY      47 ValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly----- 60

```

Db 3034817 CTCGGGAAACAGAGCCGCTGCGCTGACAGTACGGGGTCCAGGAGTACGTGCT 3034876
QY 60
Db 3034877 CCGCGCGCGCGCATGATCATGCTCCAGCACCGGCGTCCAGAGCGGCTATTACCTGCG 3034936
QY 61 ArgSerThrArgGluLeuVal-GlyAspValLeuArgAspThrLeuSerGlu----- 78
Db 3034937 CGTAGCCAGCGCGCGTACGCGCGGTGACCCCTCGCCAGACCATTCGAGCAGAGTGC 3034996
QY 78
Db 3034997 GTACCTTTCAACAGCGCCCTGATCAACATCAGTGGTGCAGCGGCCAACCGCTCGT 3035056
QY 79
Db 3035057 CGAGAAATCGTTGCTGTGACAGTCCAGCCGACGATTAACCGT-----GTCCACCC 3035104
QY 91 ogGluValThrAsp-----GlyLeuGlySer-----GlyAspGluValSlySme 105
Db 3035105 GCGCGCAGCAGATGCTGCGCGCGACCGGCAATGGCTGCCCGGTGAGTTCCAGCACT 3035164
QY 105 tGluArg-----ValSerGluProCysAspProGlySerGlySgl 118
Db 3035165 CGAGCGCCCGCACCGTCCGCGCGCATCTGCGGCGAGATTGCCGCGCTGCAGCCGGGT 3035224
QY 118 n-----LeuValPheGluGluLeuGlnCysAspVal-SerMet----- 130
Db 3035225 CTCGGCGGGTAATCTCGGCGGTGACGCGGATCAACCGGACCTCGCGACCGCGCTCCC 3035284
QY 131
Db 3035285 ATCTCGCGCTACTGAAGATCGGTGCGGAAACGACAGTCTGACGAGGATTTGGA 3035344
QY 139 hrPheThrLeuValArgPheAspValSglValValThrArgGluAspThrLeuSerL 159
Db 3035345 CATACAGCATGAGCCGCCACCAACGACCGGCTCCCGGGCTGCATCCCTCGA 3035404
QY 159 eu----- 159
Db 3035405 TGTCGCGCGCGCGCGCGCTGTAGCGCGCACGCTCGCGTTTCGGTGATCCAGGA 3035464
QY 160
Db 3035465 CATCGAGTTATGATGCGGAGATGCGACGCGCTGCTGACGCGGAGCTTCGCGTCCCGA 3035524
QY 168 spSerSerValAsnHisSerProThrSerSerGlySmeLeuArgValLeuThrValA 188
Db 3035525 TGTCATGCGCGCATACAGCTGACATCGCTGCGGCTTCAGATCTCAACGCTCACTCG 3035584
QY 188 laProAspGlySerGlnSerLysArgSer-----V 198
Db 3035585 CGCGGACCGCGCGCGACGCGAGCGCAACTGGGACTTGCAGCGCGCTCGGTCCGA 3035644
QY 198 alleValAspGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThr----- 214
Db 3035645 TAAATGCCAAC-----GTTCTACGGCTGCCAGACCGCGGAATATCCACGCGGTG 3035698
QY 215
Db 3035699 TGGGCTCTCTCGTGAAGACTCTTGGCGCATCGCGGTCCTGGCTCGGCGACGCGGCCAG 3035758
QY 217 hrGluAspLeu-----A 221
Db 3035759 CACCTGAATGCAACCGCGCAACACTGTGCGGAGACGCGGTTCAAGACGCGAGGTC 3035818
QY 221 rgSerTrpGluValSlyGlnArgAlaProLeu-----ArgPheGlnGlyAspSerA 238
Db 3035819 GCGCGTGCAGCAACGCTGCTGAAGAGCCGCTGATACCGGCGCTGCGGGGTATAGGCC 3035878
QY 238 rgLeuGluGln-----SerGlyCysTrpTrpHisCysValAspGluValIleGluArgA 256

Db 3035879 GCGGAGAGCGCGGCGCTCAGGCTGT-----TCAGCCGCTC 3035914
QY 256 rgAsnHisTrpLeuAspLeuAlaGlyIleGluValnTrpThrSerGlnPheGlyProGlyS 276
Db 3035915 GCGGAGACATGATACC-----CGATGATCGGCTCCGCTC 3035953
QY 276 erProSerValAlaGlnLysSerGluLeuProProArgThrSerAsnProThrArgSerA 296
Db 3035954 GATGTGCGCTCGAGATTCCTGCGCAGTGCACCGGCGCATCGGAAAC-----CGTGTCT 3036007
QY 296 rgSerHisGluProGluAlaIleHisIleProHisArgLysPro----- 310
Db 3036008 GCTGCATAGACCTGACCTGCTGCTGCTGCTGACGCGCGGCTGCGCTCCAGGC 3036067
QY 311
Db 3036068 GGTAACTAGCGCACACAGGGGTAAATTCACCGGAACGCACTCCGTACCGTCTTCGG 3036127
QY 326 laValSerGluLeuGlnArgLeuArgGlyThrGlnAsp-----G 341
Db 3036128 CCGCA-----GCCGACCTCGCAGCTCGGCGCGGAAGCCGCAAGGTGCCAGCTCGTGGG 3036184
QY 341 LysTrpHisIlePheValArgSerProValAlaGlnGlyLysSerValGlyValGlyHisV 361
Db 3036185 GCGCAGCAGCTGCTGCGCGCGCGCTCCGACGCAATCAGCTTT----- 3036231
QY 361 alAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProS 381
Db 3036232 -----TCGGCAACAGAGGCGCGCGGATCACCG-----CGGCTCC 3036268
QY 381 erAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysL 401
Db 3036269 CAATCGCGCAGCTCGCGACGCGCGCTCCGCGCATGTCCGGAACCAACCGCGCGGGA 3036328
QY 401 yHisLysHisIleArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuA 421
Db 3036329 CGGAACGATCCGATGCGCTCAACACACAAAGCTTAACGCTTG----- 3036375
QY 421 laSerGlyGly-----Prov 426
Db 3036376 -CGCGCGGATTTCCGCTCTGTGAGCACGCGTCCAAAGACCGTCTGAGAGCGCGA 3036433
QY 426 alleGluArgGlnHisLeuArgGluLeuProAlaLeuValValTrpGluSerGlnAlaG 446
Db 3036434 CACTGTCCCGCGCTCATTCGCGAG-----AGGTCAACGAACCGCAACTG 3036481
QY 446 ly-----GlnProValGlnArgHisGlu-HisHisHis----- 457
Db 3036482 GCCCGCGCGCCCAATTCGCGCGGAAACCGTCAATCATCGGACCTGTACGCGGCAA 3036541
QY 458 -----HisGluHisHisHisLeuTy 464
Db 3036542 GGCTACGCTCGCGCAACACACACTAC 3036571
RESULT 13
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: PRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA

```

/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 297 Length: 4411529
Percent Similarity: 149.50 Matches: 133
Best Local Similarity: 29.76% Conservative: 67
Query Match: 19.79% Mismatches: 200
DB: 5.98% Indels: 272
Gaps: 30

US-09-993-966-7 (1-470) x US-09-103-840A-1 (1-4411529)

QY 9 A1A1A1A1CyslybarghgluSerProgluGlyAspSerPheAlaValSerAla1a 28
DB 3039833 GCGGCGCATACCCGACCGGCTCTGTCGACCA----- 3039862
QY 29 TrpAlaArglybgllylelgluGluTrpLleGlyArglnArg-----CysProgluGly 46
DB 3039863 ---GTCGTGCGCGGTCTCGGTGAGCTGGCCACACCGGGGCTGAGCCAGTGCACCGGTGT 3039919
QY 47 ValSerGlyProArglnLeuArglnAglyThrLleGly----- 60
DB 3039920 CTCGGGGAACCAAGAACCGGTGCGGTGCAAGTACGGCGGGTCCAGGTACGTCTGT 3039979
QY 60 ----- 60
DB 3039980 CCGCGCGCGCGCATCATGTCTGACCTTCAGACACCGCGCTCCAGACGGCTATTACTGCG 3040039
QY 61 ArgSerThrArglnLeuVal-GlyAspValLeuArgAspThrLeuSerGluGlu----- 78
DB 3040040 CGTAGCCAGCGCGCGGTGACCGGTGACCCCTGCGCAGACCATTCGAGCAGAGTGC 3040099
QY 78 ----- 78
DB 3040100 GTACTCTTCAACACAGGCCCTGATCAACATCAGTGTGTCGACGCGGCCAACCGCTCTGT 3040159
QY 79 -----GluGluAspAspPheArgLeuGluValAlaLeuProPr 91
DB 3040160 CGAGATCGTGTCTGTGACAGTCCACCCGACCGACGATACCGT-----GTCCACCC 3040207
QY 91 oglyuLyThrAsp-----GlyLeuGlySer-----GlyAspGluLybysme 105
DB 3040208 GCGGCGCATCGATGCTGCGCGCGGACGCGCAAAATGCTGCTCCGCTGATTCACCACT 3040267
QY 105 rgluArg-----ValSerGluProCysProGlySerLybysgl 118
DB 3040268 CAGGCGCGCGCACCGTGGCGCGGCTGTGTGGGCAAGATTGCGCGGCTGCAAGCGGCT 3040327
QY 118 n-----LeuLybPheGluGluGlnCysAspVal-SerMet----- 130
DB 3040328 CTCGCGCGGCTAATCTGCGGTGACGCGCACCGGTCCGCGGTTCGATTCACGGA 3040387
QY 131 -----GluGluAspSerArglnGluTrp 139
DB 3040388 ATCTCGCGGCTAGTAAGATCGGTGGGAAACACACAGCTGTCGAGCAGGAGATTGA 3040447
QY 139 hrPheThrLeuTyraAspPheAspAsnGlyLybValThrArgGluAspLleThrSerL 159
DB 3040448 CATACAGCATTCAGCGCGGCCAACACAGACCGGACCGCTCCCGGGCTGCAGTCCCTCA 3040507
QY 159 eu----- 159
DB 3040508 TGTCCCGCGCGCGCGCTGCTGATGCGCGCACCGGTCCGCTTCGATGATCCAGGA 3040567
QY 160 -----LeuHisThrLleTyrgluValVala 168
DB 3040568 CATTCAGTTGATGATCGGAGATCCACGCGGCTGCTGACGAGGACAGTCCCGCA 3040627
QY 168 apSerSerValAsnHisSerProThrSerSerLybMetLeuArgValLybLeuThrVala 188
DB 3040628 TGTTCATGCGCGGATACAGCTGATGCGGTGCGCGCTTCACGATCTCACGCTCACTCTGG 3040687

```

```

QY 188 laProAspGlySerGlnSerLybArgSer-----V 198
DB 3040688 CCGCGACCGCGCGCGCATCGACCGCACTGGAGCTTGGCGCGCCCGCTCGATCCGA 3040747
QY 198 alLeuValaAnglnAlaAspLeuGlnSerLlaArgProArgAlaGluThr----- 214
DB 3040748 TAATCGCAAC-----GTTCTACGGCTGTCAGACACCGGCAAAATACCCACCGCTG 3040801
QY 215 -----LybProT 217
DB 3040802 TGGCGTCTCTCGGAGAACTCTTGGCCGATTCGCGGCTCTGGCTCGGCGACGCGCGCAG 3040861
QY 217 hrGluAspLeu-----A 221
DB 3040862 CACCTGAATATGCAACCGGCCAACACCTGTGCGGACCGCGGTCAAGACGCGAGGTC 3040921
QY 221 rgsertTrpGluLybysGlnArgAlaProLeu-----ArgPheGlnGlyAspSerA 238
DB 3040922 GCGGCTGGCCAGCGGTGTCGAGAGCCCGCTGCATACCGGCGCGGTGCGGTATAGCC 3040981
QY 238 rgluGluGln-----SerGlyCysTyriHisLybCysValAspGluAsn1leGluArgA 256
DB 3040982 GCGGAGAGCGCGGGGCGGTCAAGGTGT-----TCAGCGCTC 3041017
QY 256 rgaAsnHisTyriLeuAspLeuAlaGlyLleGluAsnTyriThrSerGlnPheGlyProGlyS 276
DB 3041018 GCGGACGACATGATGACC-----CGATCGGATCGGAGCTCCCGGTC 3041056
QY 276 expProSerValAlaGlnLybSerGluLeuProProArgThrSerAspProThrArgSerA 296
DB 3041057 GATGTGCGCTGCGCATGTTGCTCTGCCACGTGCCACCGGCGCATTCGAGAC-----CGTGGTC 3041110
QY 296 rgsertAspGluProGluAlaLleHisLleProHisArgLybPro----- 310
DB 3041111 GCTGGCATTAACGTGAGACGTGTCCTGCTCAGGCGCGGCGCTGCGCCGTAACCGAGCC 3041170
QY 311 -----GlnGlyValAspProAlaSerPheHisPheLeuAspThrProLleA 326
DB 3041171 GTTAAGTAGCGCACACAGGGGTATTCACCGGACGCGCAACTCCGTACCGGTCTGCGG 3041230
QY 326 laLybValSerGluLeuGlnArgLeuArgGlyThrGlnAsp-----G 341
DB 3041231 CCGCA-----GCCGACCTCGACAGTGGCGCGGACCGCGCAAGGTGCCAGTGGTGG 3041287
QY 341 lySerLybHisPheValArgSerProLybAlaGlnGlyLybSerValGlyValGlyHisV 361
DB 3041288 GCGCAGACGCTCTGCGCGCGCGGCTCCGACAGCAATCAGCTT----- 3041334
QY 361 alAlaArgGlyAlaArgAsnLybProProLeuGlyProAlaLleProAlaValSerProS 381
DB 3041335 -----TCGGCAACAAAGAGCGCGCGGATCACCG-----CGGCCC 3041371
QY 381 erAlaHisLybMetAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLybL 401
DB 3041372 CAATCGCGCGCATCTCGGACGCGGCGGCTCCGCGCACTTCGGAACCAACACCGCGCGGA 3041431
QY 401 ybHisLybHisArgAlaLybGluSerGlnGlnGlyCysArgGlyLybGlnAlaProLeuA 421
DB 3041432 CGGAGACATCCGATGGCGCTCAACACAAACAAAGTAAAGCTTG----- 3041478
QY 421 laSerGlyLyb-----Prov 426
DB 3041479 --GCGGCGGATTCGCGCTCTGTGAGCAACCGCTGCAAGAGACCGTGTGAGAGCGCGA 3041536
QY 426 alLeuGluLybArgGluHisLybLeuArgGluLeuProAlaLeuValTyriGluSerGlnAlaG 446
DB 3041537 CACTGTCCGCGCTCATTCGCGCAGAG-----ACGATCAACGAACCGCAAGCTG 3041584
QY 446 ly-----GlnProValGlnArgHisGlu-HisHisHisV----- 457
DB 3041585 GCGCGGCGCGCCCAATTCGCGCGGAGAACCGGTGATCATTCGCAACCTCGTACGCGCGCA 3041644

```

QY 458 -----HisGluHisHisHisTyr 464
 Db 3041645 GGCTACGTCGACACACACACTAC 3041674

RESULT 14
 US-08-471-033-35
 Sequence 35: Application US/08471033
 Patent No. 5770696
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Koziele, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 City: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,033
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: P-40,403
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 35:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2576 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "Synthetic DNA"
 HYPOTHEICAL: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..2564
 OTHER INFORMATION: /note= "Maize optimized sequence
 OTHER INFORMATION: encoding VP1a(a) with the Bacillus secretion signal removed
 OTHER INFORMATION: contained in pCIB552c"
 US-08-471-033-35

Alignment Scores:
 Pred. No.: 0.00728 Length: 2576
 Score: 146.50 Matches: 101
 Percent Similarity: 33.33% Conservative: 36

Best Local Similarity: 24.57% Mismatches: 115
 Query Match: 5.86% Indels: 160
 DB: 1 Gaps: 21

US-09-993-966-7 (1-470) x US-08-471-033-35 (1-2576)

QY 190 AspGlySerGlnSerLysArgSerValLeuValAsnGlnAlaAspLeuGlnSer----- 207
 Db 56 GATGACCCGACAGGCGCTGCTGGCTACTACTTCAAGGCGAGGACCTTCAACCTGAC 115
 QY 208 ---AlaArgProArgAlaGlnThrLysProThrGlnAspLeuArgSerTrpGluLys 226
 Db 116 CATGTCGCGCCCGACCGTACAGACACCTT-----GATCAGACACGACGCGCCAA 169
 QY 227 GlnArgAla-----ProLeuArg 232
 Db 170 CAAGCTGCTGACAGAGAGAGAGAGATGACAGAGATCCGCTGATCGGCTGATCCA 229
 QY 233 PheGlnGlyAspSerArgLeuGlu----- 240
 Db 230 GAGCAAGAGACCGGCGACTTCACTTCAACCTGAGGAGAGAGAGGCGCATCATCA 289
 QY 240 ----- 240
 Db 290 GATCAACGCGACATCATCAACAAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAA 349
 QY 241 ---GlnSerGlyCysTrpThrHisLysCysValAspGlnAsnLglnLysArgAsnHisLys 259
 Db 350 GGGCAAGCTGCTGCTCCCATCAAGATCGAGATCCAGAGACACCAAGTTCAACATCGACAG 409
 QY 260 LeuAspLeuAlaGlyLglnGluAsnTrp----- 268
 Db 410 CAAGACCTTCAAGAGAGTGAAGCTTTCAAGATCGACAGCAGACAGACAGCCCGAGAGGT 469
 QY 269 -----ThrSerGlnPheGly---ProGlySerProSer 278
 Db 470 GCACAGAGACGAGCTGGCAACCCCGAGTTCAACAGAGAGAGAGAGAGAGAGTTCCTGGC 529
 QY 279 ValAlaGlnLysSerGluLeuProProArgThrSerAsnProThrArgSer---ArgSer 297
 Db 530 CAAGCCGACGACATCA---CCTGTTACCCAGCATGAAGCGGACGATTCGACGAGGA 586
 QY 298 HisGlnProGlnAlaLeuHisLysProHisArgLysProGlnGlyValAspProLysSer 317
 Db 587 CACCGA-----CACCGAGCGGACAGCATCCCGACCTGTGGGA 625
 QY 318 PheHisPheLeuAspThrProLeuAlaLysValSerGluLeuGlnAlaGlyLeuArgLys 337
 Db 626 GGAAGACGGCTACACATCA-----GAAACCGATCCCGTGAAGTGGGA 670
 QY 338 ThrGlnAspGlySerLys-----HisPheValArgSerProLysAlaGlnGly 354
 Db 671 CGACAGCTGCTGACAGGCGCTACACCAAGTTCTGTAGCAACCC---CTGGAGACCA 727
 QY 354 sSerValGlyValGlyHisVal-----AlaArgGlyAlaArgAsnLysPro----- 369
 Db 728 CACCGTGGGCGACCCCTCAACACGACGACGAGAGAGCGCCCGGACCTGACCTGACCAA 787
 QY 370 -----ProLeuGlyProAlaLglnProAlaVal----- 378
 Db 788 CGCAAGAGACCTTCAACCCCTGTGGCGGCTTCCCGAGGTGAACCTGAGCATGGA 847
 QY 379 -----SerProSerAlaHisLeuAlaLysProLysAlaLeuLeuProSerLe 394
 Db 848 GAAGGTATCTGAGCCCAAG 892
 QY 394 uAlaProLeuGlyHisLysLysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArg 414
 Db 893 -GAGCCACTGAGACCACTGAGACGACACCAACAGAGAGAGAGAGAGAGAGAGAGAGAG 951
 QY 414 gGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlnHisLysLeuArgL 434
 Db 952 GCAT-----CGTCCCAAGGAGCATCACTTCGCGGTAG-----CGTGA 990


```
QY 370 -----ProLeuGlyProAlaIleProAlaVal----- 378
Db 788 CGCCAGAGAGACCTTCAACCCCTGCGTGGCCCTTCCCGAGCGTGAACTGAGCATGGA 847
QY 379 -----SerProSerAlaHisLeuAlaIleSerProAlaLeuLeuProSerLe 394
Db 848 GAAGGTGATCTTGAGCCCAACGAACTGAGCAACAGCGT-GGA----- 892
QY 394 uAlaProLeuGlyHisIleLysHisIleArgAlaLysGluSerGlnGlnGlyCysAr 414
Db 893 -GAGCCACTCGAGCACCACCTGAGCTACACCAACCCGAGGCGCCAGCGTGAGGCCG 951
QY 414 gGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlyHisLeuArgG 434
Db 952 GCAT-----CGGTCCAGAGGACATCAGCTTCGGCGTGAG-----CGTGA 990
QY 434 uLeuProAlaLeu-----ValValTyrGluSerGlnAlaGlyGlnPr 448
Db 991 ACTACCAAGCACAAGGAGACCGGTGGCCAGAGAGTGGGGCACCAAGCAGCAACCAAGCC 1050
QY 448 oValGln----- 450
Db 1051 AGTTCAACACCGCAGCGCGCTAAGCTGAACGCCAAGTGGCTGACACACAGTGGGCA 1110
QY 451 -----ArgHisGlyHisHisHis-----HisHisGly 459
Db 1111 CCGGCGCCATCTAGAGCGTAAGCCCAACCAAGCTTCGTGCTGAACAGACGACCATCG 1170
QY 459 uHisHisHisHisIleTyrHisHisPheTyrGln 469
Db 1171 CCACCATCACCGCAAGTCGATTCACCGC 1201
```

Search completed: December 30, 2004, 09:00:02
Job time : 3940 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: December 30, 2004, 03:20:04 ; Search time 5644 Seconds
(without alignments)
3034.492 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 1 MGJLHSPAAVCKRRSPSG.....RHHHHHHHHHHHPYQT 470
Sequence: 1 MGJLHSPAAVCKRRSPSG.....RHHHHHHHHHHHPYQT 470

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xip
-O/cgmt 1/USPTO.spool.p/US09993966/runat.29122004.130758.9009/app.query.fasta.1.647
-DB=EST -OPMT=fastap -SUFFIX=rat -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blodum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09993966 @CGM 1.1 6425 @runat.29122004.130758.9009 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJOURN -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -XLONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2489	99.6	2142	3	AF289584	Homo sapi
2	2179.5	87.2	2289	3	AK082367	Mus muscu
3	2094.5	83.8	2471	3	AK076164	Mus muscu
4	1852.5	74.1	1466	3	AK013241	Mus muscu
5	1545	61.8	885	9	AY412098	Homo sapi
6	1462	58.5	931	5	B0653673	AGENCOURT
7	1436	57.5	921	5	B0645656	AGENCOURT
8	1420	56.8	990	5	B064678	AGENCOURT
9	1397	55.9	895	5	B0644360	AGENCOURT

10	1382	55.3	928	5	B0644956	AGENCOURT
11	1374	55.0	888	9	AY412100	Mus muscu
12	1283.5	51.4	1008	5	B0646371	AGENCOURT
13	1252	50.1	996	5	B0652087	AGENCOURT
14	1240	49.6	885	9	AY412099	Par trogl
15	1216	48.7	966	5	B0649813	AGENCOURT
16	1195.5	47.8	915	6	CB182455	AGENCOURT
17	1186.5	47.5	1027	4	BG104777	AGENCOURT
18	1118	44.7	899	5	B0645507	AGENCOURT
19	1092	43.7	913	5	B0895164	AGENCOURT
20	1078.5	43.2	953	5	B0846054	AGENCOURT
21	1063	42.4	674	6	CB056689	NISC j119
22	1060	42.4	677	4	CB731367	UT-M-G20
23	1031	41.3	627	7	B1767278	UT-M-G20
24	1028	41.1	777	4	BG542261	UT-M-G20
25	1021	40.9	724	6	CA328380	UT-M-PYO
26	1000.5	40.0	946	6	BY712252	BY712252
27	978	39.1	839	7	CK795599	CK795599
28	975	39.0	1059	5	BQ070932	AGENCOURT
29	974	39.0	564	4	BM711145	UT-E-DX1
30	918	36.7	653	2	BF540602	BF540602
31	911	36.5	678	2	BF537335	BF537335
32	897.5	35.9	825	6	CB961961	AGENCOURT
33	885	35.4	934	2	BF167269	BF167269
34	867.5	34.7	2399	3	AK008814	Mus muscu
35	857	34.3	751	4	BG820139	AGENCOURT
36	855	34.2	622	6	CB215756	CB215756
37	839.5	33.6	600	4	B1990034	4054-43 M
38	830	33.2	540	5	BX513216	BX513216
39	818.5	32.8	693	5	BM935129	UT-M-BH3
40	811	32.5	670	2	BB650914	BB650914
41	810	32.4	2802	3	AK081904	Mus muscu
42	808.5	32.4	684	7	CK450648	CK450648
43	803	32.1	481	4	B1047069	MR3-FN020
44	797.5	31.9	826	6	CB961668	CB961668
45	793	31.7	522	4	B1903274	id24d04.y

ALIGNMENTS

RESULT 1	AF289584	2142 bp	mRNA	linear	HTC 01-JAN-2002
LOCUS	AF289584				
DEFINITION	Homo sapiens clone pp7246 unknown mRNA.				
ACCESSION	AF289584.1	GI:18027371			
VERSION					
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Man,D.F. and Gu,J.R.				
TITLE	Novel human cDNA clones with function of inhibiting cancer cell growth				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Man,D.F. and Gu,J.R.				
AUTHORS	Direct Submission				
TITLE	Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/ln 2200 Xie-Tu Road, Shanghai 200032, P. R. China				
JOURNAL					
FEATURES	Location/Qualifiers				
SOURCE	1..2142				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="pp7246"				
	277..1689				
	/codon_start=1				
	/product="unknown"				
CDS					

```

/protein_id="AAU55768.1"
/db_xref="GI:18027372"
/translation="WGLKHSKFDVCKRRSPBGDSPAVSANARKIEIEMWIGRQCP
GVGSPQLRLAGTIGRSTRELVDVLRDLSEEDFLREVALPEKIDGLSGDE
KKMERVSEPCPSPSKQKLFELQCDVAMEESRQEMFTVLDPNNKQKVRDETTSL
HTIYEVDSSVNSPTSSKMLRVKLTVAPDOSKRSVLVNOALQGARPAETKPTPE
DLSEWKKORAPLARFOGDSRLFOGSHHCNDENIERRNYHLDLAGIENTSOQPS
PSVAOKSELPRPTNSPTRRSRSHBEPAIHITRRKQGVDPASFTIIDLPIAKVSELOQR
LRGTQDSKHPVSPKAQGSVGVHVARGRNRPGLGPAIPAVSPAHILASPALPVR
SLAPLGHKHKHRAKRSQQCGRLQAPLASGVPVLRHRLRELPALVVESQAGPQV
RHEHHHHHHHHHHFYQT"

```

ORIGIN

Alignment Scores:

```

Pred. No.:      5.6e-191      Length:      2142
Score:          2489.00      Matches:      468
Percent Similarity: 99.57%      Conservative: 0
Best Local Similarity: 99.57%      Mismatches: 2
Query Match:    99.60%      Indels:      0
DB:             3             Gaps:      0

```

US-09-993-966-7 (1-470) x AF289584 (1-2142)

```

Qy      1  MetGlyLeuHisSerLysProAlaAlaValCysLysArgArgGlySerProGlyGly 20
Db      277 ATGGGAAACTTCTCTCCAGCCCGTGCAGCTGTGCAGAGCCGAGAGAGCCGGAAGCT 336
Qy      21  AspSerPheAlaValSerAlaAlaTAPAlaArgLysGlyIleGlyIleTyrIleGlyArg 40
Db      337 GACAGCTTCGCGCGTGCAGCGCTGCTGGAGGCGCATCGAGAGTGTGATCGGAGGA 396
Qy      41  GlyArgCysProGlyGlyValSerGlyProArgGlyIleuArgLysAlaGlyTyrIleGly 60
Db      397 CAGGCTGCGCCGGCGGCTGTCTGGAGCCCGAGACGCTGCGTGGCGGAGCAATGAGC 456
Qy      61  ArgSerThrArgGlyLeuValGlyAspValLeuArgSerThrLeuSerGlyGlyGlyGly 80
Db      457 CGAAGCACCCCGGAGCTGTGGGCGACCTGTGAGAGACAGCTCAGAGAGAGAGAGAG 516
Qy      81  AspArgPheArgLeuGlyValAlaLeuProProGlyLysThrAspGlyLeuGlySerGly 100
Db      517 GACACCTTTCGCGTGCAGAGTGGCCCTGCTCTCTGTAAGAGATGACGCGGCTGGAGCGGA 576
Qy      101  AspGlyLysLysMetGlyArgValSerGlyProCysProGlySerLysGlyLeuLys 120
Db      577 GATGAGAAAGATGAGAGAGTAGAGCAACCTCTGCCACGCTCCAGAGACACTGTAAG 636
Qy      121  PheGlyGlyLeuGlyCysAspValSerMetGlyLysAspSerArgGlyIleTyrThrPhe 140
Db      637 TTGTAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGAGACCGCGAGAGTGAACCTTC 696
Qy      141  ThrLeuTyrAspPheAspAsnAsnGlyLysValThrArgGlyAspIleThrSerLeuLeu 160
Db      697 ACCCTGATGACTTTGACCAACACGCGCAAGGTCAACCCGAGAGACATCAACAGCTTGCTG 756
Qy      161  HisThrIleTyrGlyValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db      757 CACACCACTATGAGGTGAGCTCTCTGTCAACCACTCCCAACACTTCAGCAAGATG 816
Qy      181  LeuArgValLysLeuThrValAlaProAspGlySerGlyIleSerLysArgSerValLeuVal 200
Db      817 CTGGGGGTAAGCTCACCGTGGCCCGATGCGACGACGAGCAAGAGAGAGCTCTTGTTC 876
Qy      201  AsnGlyAlaAspLeuGlyIleSerAlaArgProArgAlaGlyIleThrLysProThrGlyAspLeu 220
Db      877 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGAGACAGACCAAGCCCATGAGAGACTG 936
Qy      221  ArgSerThrGlyLysLysGlyArgAlaProLeuArgPheGlyIleLysAspSerArgLeuGly 240
Db      937 CGGAGCTGGGAGAGAGACGAGCGCCGCTCAGGTTCCAGGGGTGAGACGCGCTGAGAG 996
Qy      241  GlySerGlyCysTyrHisHisCysValAspGlyLysAsnIleGlyArgArgAsnHisTyrLeu 260

```

```

Db      997 CAGTCTGGCTGTACCAACCATGTGGATGAGAACATCGAGAGAGAAACCACTACTTA 1056
Qy      261  AspLeuAlaGlyIleGlyValIleTyrThrSerGlyPheGlyLysSerProSerValAla 280
Db      1057 GATCTCCCGCGGATAGAAACCTACAGCTCCCAATTTGGGCTGTGCTCCCTTCGTVGCC 1116
Qy      281  GlyLysSerGlyLeuProProArgGlyThrSerAsnProThrArgSerArgSerHisGlyPro 300
Db      1117 CAGAGTGTCAAACTGCCCCCCCCCGACCTCCAAATCCCACTGATCTCGCTCCCATGAGCCG 1176
Qy      301  GlyAlaIleHisIleProHisArgLysProGlyIleValAspProAlaSerPheHisPhe 320
Db      1177 GAAGCATTCACATCCACACCGAAGCCCGAAGCGTGAACCGGCTCTTCCACTTC 1236
Qy      321  LeuAspThrProIleAlaLysValSerGlyLeuGlyIleArgLysLeuArgGlyTyrThrGlyAsp 340
Db      1237 CTTGACACCCCAATCGCCAGAGTCTCAGAGCTCCAGCAGCGCTCCGCGGACCCAGAGAC 1296
Qy      341  GlySerLysHisPheValArgSerProLysAlaGlyIleLysSerValGlyValGlyHis 360
Db      1297 GGGAGCAGACCTTTGTGAGGTCCCCCAGGCCCCAGGCGCAAGAGTGTGGGTGGGCCAC 1356
Qy      361  ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db      1357 GTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAGCCCGCCATCCCTGCGGTGTCCCCC 1416
Qy      381  SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1417 TCGGCCACCTGGCTGTCACGCCCGGCTCTCTCCCTCTTACGCCCTTCGGGACAAAG 1476
Qy      401  LysHisLysHisArgAlaLysGlySerGlyIleGlyCysArgGlyLeuGlyAlaProLeu 420
Db      1477 AAGCACAAGACCCAGGACAGAGAGAGCAGCAGGCTGCCGGGCTCGAGGACCACTG 1536
Qy      421  AlaSerGlyGlyProValLeuGlyArgGlyHisLeuArgGlyLeuProAlaLeuValVal 440
Db      1537 GCCTCAGAGTGGCTGTGCTGTGGGCGGAGCACTGCGGAGGCTGCCGCTGTGTGTG 1596
Qy      441  TyrGlySerGlyAlaGlyIleProValGlyIleArgHisGlyHisHisHisHisHisHis 460
Db      1597 TATGAGAGCAGGCGCGGCGAGCGGTCCAGAGCATGAGACCAACCAACCATGAACAT 1656
Qy      461  HisHisHisTyrHisHisPheTyrGlyIleThr 470
Db      1657 CACCACCATTAACCACTTCTTACAGACA 1686

```

RESULT 2
AK082367
LOCUS
DEFINITION
MUS MUSCULUS 0 DAY NEONATE CEREBELLUM CDNA, RIKEN FULL-LENGTH
(DROSOPHILA), FULL INSERT SEQUENCE.

ACCESSION
AK082367
VERSION
AK082367.1 GI:26100612
KEYWORDS
HTC; CAP TRAPPER.
SOURCE
MUS MUSCULUS (HOUSE MOUSE)
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
MEDLINE
PUBMED
99279253
10349636
2

AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K.,
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
20530913
11076861

REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE
AUTHORS
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6 (bases 1 to 2289)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imelani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numata, K., Ono, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
location/Qualifiers

FEATURES
source
1..2289
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C230040009"
/db_xref="taxon:10090"
/clone="C230040009"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
152..1568
/note="naked cuticle 1 homolog (Drosophila)"
(MGI:2135954, GB|NM_027280, evidence: BLASTN, 98%,
match=1636)
misc_feature
match=1636

ORIGIN
Alignment Scores:
Pred. No.: 7.73e-166 Length: 2289

Score: 2179.50 Matches: 410
Percent Similarity: 90.87% Conservative: 18
Best Local Similarity: 87.05% Mismatches: 41
Query Match: 87.21% Indels: 2
DB: 3 Gaps: 1
US-09-993-966-7 (1-470) x AK082367 (1-2289)

QY 1 MetGlyLysLeuHisSerIlyPProAlaAlaValCylValArgArgGlnSerProGlnGly 20
DB 152 ATGGGGAACTTCACTCAAGAGCCGGCCGGTGGCAAGGCGAGAGAGAGAGAGAGT 211

QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyLleGlnGlnTTPllleGlyArg 40
DB 212 GACAGCTTTCCTTAAGCGCTGCTGGCAAGAAAGGCAATCAAGAGTGTATCGGAGG 271

QY 41 GlnArgCysProGlnGlyValLserGlyProArgGlnLeuArgLeuAlaGlyTTrlleGly 60
DB 272 CAGCGCTGTCAGGACAGCGCTCTCAAGACCCCGTCAGCTGAGATTGGCAGGACTTGGT 331

QY 61 ArgSerThrArgGlnLeuValGlyAspValLleuArgAspThrLeuSerGlnGlnGln 80
DB 332 CGAGGCACTCGGAACTCGTGGGTGACACTTCAAGAGAGGCTCGGTGAGAGAGAG 391

QY 81 AspAspPheArgLeuGlnValAlaLeuProProGlnLysThrArgGlyLeuGlySerGly 100
DB 392 GACGACTTCCCTTAGAAGTGGCCCTGCGCCGCAAGAAATTAACAGCTAAGTATGGA 451

QY 101 AspGlnLysLysMetGlnArgValLserGlnProCysPheProGlySerLysGlnLeuLys 120
DB 452 GATGAGAAAGAAATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511

QY 121 PheGlnGlnLeuGlnCysAspValLserMetGlnLysAspSerArg-GlnGlnTTPTrp 140
DB 512 TTGGAAGAGCTGAGTGTATGTCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 571

QY 140 eTrileuTyrrAspPheAspAsnAnglyLysValThrArgGlnAspLleThrSerLeu 160
DB 572 CACTTATATGACTTGCACAACATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631

QY 160 uHsTTrlleTylGlyValValAspSerSerValAsnHisSerProThrSerSerLys 180
DB 632 GCATTCACATTTAAGAGTGTATGTCTCTGTGAGAGAGAGAGAGAGAGAGAGAG 691

QY 180 tLeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeu 200
DB 692 ACTGGCGGTAACTCAACCGTGGCTCTGAGGAGAGAGAGAGAGAGAGAGAGAG 751

QY 200 laeGlnAlaAspLeuGlnSerAlaArgProArgAlaGlyThrLysProThrGlnAsp 220
DB 752 CAACCATTCAGATCTGAGAGCAAGAGCCCGCAGAGAGAGAGAGAGAGAGAG 811

QY 220 uArgSerTrpGlnLysLeuGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeu 240
DB 812 GCGTGGCTGGGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871

QY 240 uGlnSerGlyCysTyrTrlHisCysValAspGlnAsnLleGlnLysArgAsnHisTyrLe 260
DB 872 GCAGCCAGAGATGCTACCATTTGGTGGATGAGAACATTGAGAGAGAGAGAGAG 931

QY 260 uAspLeuAlaGlyTllegLysAntTyrThrSerGlnPheGlyPProGlySerProSerVal 280
DB 932 AGACCTGGCGGAGATAGAAATCAACGCTTCAGTTTGACCGGAGATCCCTTCGGGG 991

QY 280 agLysSerGlnLeuProProArgTrpThrSerAspProThrArgSerArgSerHisGly 300
DB 992 CGAAGATCAAGAGCTGCCCTCGAATCTCAACCCCATCTCGCTCCCAAGAGCC 1051

QY 300 ogLysAlaIleHisIleProHisArgLysPProGlnGlyValAspProAlaSerPheHis 320
DB 1052 AGAAGCTGCGCAATCCCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1111

QY 320 eleuAspThrProIleAlaLysValSerGlnLeuGlnArgLeuArgGlyThrGlnAs 340

```

Db      1112 CTTGACACCCCATTTGCGCAAGCATCAGAGCTCAGCAACGGCTCCGGGGCATCTCAGGA 1171
Qy      340  PGIYserLyshIshPheValArgSerProLyAlaGlnGlyLyseValGlyValGlyH 360
Db      1172 TGGGAGACAGACACTTGTGTGAGGTCTCCCAAGGCCAGGGCAAGAACATGAGGTATGGGCCA 1231
Qy      360  eValAlaArgGlyValArgshnlyProProLeuGlyProAlaIleProAlaValSerPr 380
Db      1232 CGGGGCGAGAGGTGCAAGAAAGAACCTCCATCTGTATCCCAACCACTACTGTCTCC 1291
Qy      380  oSerAlaIshIshLeuAlaIshSerProAlaLeuLeuProSerLeuAlaProLeuGlyIshLy 400
Db      1292 CTCGCCCATCTGGCCACAGCCAGCCCTTCTCCCACTGGCACCCCTGGGGGCACAA 1351
Qy      400  sLyshIshLyshIshArgAlaLyshGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420
Db      1352 GAAACACAGACATCGAGCCAAAGAGACCGAGCGAGCTGCCGGGGCTGCAGGGCCCT 1411
Qy      420  uAlaSerGlyGly---ProValLeuGlyArgGlyUshIshLeuArgGlyLeuProAlaLeuVa 439
Db      1412 GCGTCGAGAGGCTCCACCGTCATGAGGAGCGGAGACAGGTGAGGAGCTGCTGCTGCTGT 1471
Qy      439  lValTyGlySerGlnAlaGlyGlnProValGlnArgHiseGlyUshIshIshIshIshIsgl 459
Db      1472 GGTGTACAGAGACCGAGCTGGGCGAGGCCCTCCAGAGACAGCAACCATCACCACACGA 1531
Qy      459  uHshIshIshIshIshTyHshIshIshIshIshIshIshIshIshIshIshIshIshIsh 469
Db      1532 ACATCACACCATATCACCACTCTATCAG 1562

RESULT 3
AK076164
LOCUS      2471 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
              enriched library, clone:330401D10 product:naked cuticle 1 homolog
              (Drosophila), full insert sequence.
ACCESSION  AK076164.1 GI:26096648
VERSION     AK076164.1
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1
AUTHORS     Carninci, P. and Hayashizaki, Y.
TITLE       High-efficiency full-length cDNA cloning
METH        Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     99279253
PUBMED      10349636

REFERENCE   2
AUTHORS     Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
METH        Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20499374
PUBMED      11042159

JOURNAL     3
MEDLINE     11042159
PUBMED      11042159
AUTHORS     Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
            Kono, H., Akiyama, D., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, S., Nishibe, T., Harada, A.,
            Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Watanishi, M.,
            Ozoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE       RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multipillarary sequencer
METH        Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530913
PUBMED      11076861

JOURNAL     4
MEDLINE     11076861
PUBMED      11076861
AUTHORS     The RIKEN Genome Exploration Research Group Phase II Team and the

```

```

TITLE      PANTOM Consortium.
JOURNAL     Functional annotation of a full-length mouse cDNA collection
REFERENCE   Nature 409, 685-690 (2001)
AUTHORS     5
            The PANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE      Analysis of the mouse transcriptome based on functional annotation
JOURNAL     of 60,770 full-length cDNAs
REFERENCE   Nature 420, 563-573 (2002)
AUTHORS     6 (bases 1 to 2471)
            Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
            Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
            Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
            Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
            Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
            Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
            Nakamura, M., Nishi, K., Nomura, K., Numataki, R., Ohno, M., Ohnato, N.,
            Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
            Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
            Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
            Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yaunishi, A.,
            Muramatsu, M., and Hayashizaki, Y.
TITLE      Direct Submission
JOURNAL     Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku,
            Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
            URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
            Fax: 81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL: http://genome.gsc.riken.jp/
            URL: http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
            1..2471
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM,DB:330401D10"
                /db_xref="taxon:10090"
                /clone="330401D10"
                /tissue_type="head"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="14, 17 days embryo"
                155..1571
                /note="naked cuticle 1 homolog (Drosophila)"
                (MGI|MG1:2135954, GB|NM_027280, evidence: BLASTN, 98%,
                match=1658)
                putative"

misc_feature
            1571..2471
                Alignment Scores:
                Pred. No.: 6,75e-159 Length: 2471
                Score: 2094.50 Matches: 400
                Percent Similarity: 89.41% Conservative: 22
                Best Local Similarity: 84.75% Mismatches: 47
                Query Match: 83.81% Gaps: 4
                DB: 3
            US-09-993-966-7 (1-470) x AK076164 (1-2471)
            Qy      1  MecGlyLyshIshIshSerLyProAlaIshValCysLyArgArgGlySerProGlyGly 20
            Db      155  ATGGGGAACCTTCACTGCAAGCCGCCCTGTGCAAGCGAGGAGAGAGCCCGAAGGT 214
            Qy      21  ApsSerPheAlaValSerAlaIshIshIshIshIshIshIshIshIshIshIshIshIsh 40
            Db      215  GACAGCTTCTGTAAACCGCTGCTGGGCAAGAAAGCATCGAGAGTGCAGATCGGAGG 274

```

Qy 41 GlnArgCyprGlyValSerGlyProArgGlnLeuAlaGlyThrIleGly 60
 Db 275 CAGGCGTGCAGGACGATATCAGAGCCCGCTCAGGATTCGAGCGAGCTTGGT 334
 Qy 61 ArgSerThrArgGlnLeuValGlyAspValLeuArgSerThrLeuSerGlnGluGlu 80
 Db 335 CGAGGCACTGGGAACTGTGGGTGACACTTCTACAGAGGCTTCGGTGAAGAGAGCAG 394
 Qy 81 AspArgPheArgLeuValAlaLeuProProGlnuYrThrAspGlyLeuGlySerGly 100
 Db 395 GAGACATTCCTCCCAAGAGTGGCCCTGCGCTGAGAAAGATTGACAGCTTGAAGTGA 454
 Qy 101 AspGlnuYrMetGlnArgValSerGlnuProCyprGlySerGlyValGlnuYr 120
 Db 455 GATGCAAGAGAAATGGAGAGACTGACCACTGGCCAGGCTCCAGAAAGACAGCTCAG 514
 Qy 121 PheGlnuGlnuGlnuCyprAspValSerMetGlnuArgSerArgGlnuIleThrPhe 140
 Db 515 TTGGAACAGCTGCAGGTGATGCTCTGTGAGAGAGACAGCCGCAAGAGTGAATTC 574
 Qy 141 ThrLeuYrAspPheAspAsnAsnGlyValValThrArgGlnuAspIleThrSerLeu 160
 Db 575 ACTTATATGACTTCGACAAATGGCAAGTGAACCTGAGACATTACAGCTTGTCTG 634
 Qy 161 HisThrIleTyrGlnuValAspSerSerValAsnHisSerProThrSerSerLeuMet 180
 Db 635 CATACCATCTATGAAGTGTGATCTCTGTGAAACCATTCCTCCACATCAAGCAACA 694
 Qy 181 LeuArgValuYr-LeuThrValAlaProAspGlySerGlnuSerIleYrArgSerVal 200
 Db 695 CTGGGGGTGAAGCCTCAACCGTCTCTGAGCGGAGCCAGTGAAGAGAGACGCTCTTT 754
 Qy 200 IaenGlnAlaAspLeuGlnuSerAlaArgProArgAlaGluThrIleYrProThrGlu 220
 Db 755 CAACCATACCGATTCGAGAGCAACAAGCCCCGGCGACA-ACCAACCCCGTGAAGAGCT 813
 Qy 220 uArgSerTyrGlnuYrGlnuArgAlaProLeuArgPheGlnuYrAspSerArgLeu 240
 Db 814 GCGGTGGTGGAGAAAAGCAGCGAGCCCTCAGGTTTCAGGGTGCACCCACCTGGA 873
 Qy 240 uGlnuSerGlyCyprThrHisGlyValAspGlnuAsnIleGlnuYrArgAsnHisGly 260
 Db 874 GCGAGCCAGACTGTACACCATTCGCTGATGAACAATGAGAGGAAACCACTACT 933
 Qy 260 uAspLeuAlaGlyIleGlnuYrThrSerGlnuPheGlyProGlySerProSerValAla 280
 Db 934 AGACCTGGCGGGATTAAGAACTACAGCTCTCAGTTTGAACGGGATCCCTTGGTGGC 993
 Qy 280 aglnuYrSerGlnuLeuProProArgThrSerAsnProThrArgSerArgSerHisGlu 300
 Db 994 CCAAGATCAGAGGTGCCCCCTCGAATCTCCAACTCCCACTGCTGCTCCACGAGCC 1053
 Qy 300 OGlnuAlaIleHisIleProHisArgGlySerGlnuYrValAspProAlaSerPheHis 320
 Db 1054 AGAAGCTGCCCACTCCACACCGAGAGCCCAAGGCTGAGACCAAGCTCTCCCACT 1113
 Qy 320 eLeuAspThrProIleAlaYrValSerGlnuLeuGlnuArgLeuArgGlyThrGln 340
 Db 1114 CTTTGAACACCCATTTGCCAGAGCATCAGAGCTCCAGCAAGCTCCGGGCACTCAGGA 1173
 Qy 340 pGlySerLeuHisPheValArgSerProYrAlaGlnuYrAspSerValGlyValGly 360
 Db 1174 TGGGAGCAACACTTTGTGAGGTCCCAAGGCCCAAGGCAAACTATGGTATGGGCA 1233
 Qy 360 eValAlaArgGlyValArgAsnYrProProLeuGlyProAlaIleProAlaValSer 380
 Db 1234 CGGGGCGAGAGGTGCAAGAAAGCAAGCTCCACTGTATCCCAACCACTATGCTCTCCC 1293
 Qy 380 oSerAlaHisLeuAlaIleAspProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
 Db 1294 CTCGCCCATCTGGCCACAGCCCAAGCCCTTCTCCCACTGGCACTCCCTGGGCGACA 1353

Qy 400 eLYeHisLeuHisArgAlaYrGlnuSerGlnuGlnuYrCyprArgGlyLeuGlnuAlaPro 420
 Db 1354 GAAACACAGCATCTGAGACCAAGAGACCCAGCGAGCTGCGGGGCTGCAAGGCCCT 1413
 Qy 420 uAlaSerGlyYr---ProValLeuGlyYrArgGlnuHisLeuArgGlnuLeu-ProAlaLeu 439
 Db 1414 GCGTCGAGAGAGCTCCACCGCTCATCGGGCGGAGCAGTGAAGAGCTGCGCGCTGG 1473
 Qy 439 aValYrYrGlnuSerGlnuAlaGlyGlnuProValGlnuArgHisGlnuHisHisG 459
 Db 1474 TGGTGTACGAGAGCGAGCTGGCGAGCGCTCCAGAGACAAACCATCACCAACAGC 1533
 Qy 459 IuHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 469
 Db 1534 AACATCACACCATTTACACCACTTATCAG 1565

RESULT 4
 AK013241
 LOCUS 1466 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
 (Drosophila), full insert sequence.
 ACCESSION AK013241.1 GI:12850482
 VERSION AK013241.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kikunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanaka, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

REFERENCE
 4 THE RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

REFERENCE
 5 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

REFERENCE
 6 (bases 1 to 1466)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiyamichi, K.,
 Hirooka, T., Hori, P., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	3,26e-108	Length:	931
Score:	1462.00	Matches:	270
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	58.50%	Indels:	0
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x B0653673 (1-931)

```

Qy 201 ASnglnAlaAspleuGInserAlaArgProArGAlaGluThrlyPProThrgluAspleu 220
Db 3 AATCAGGCTGACTGCGAGAGCGCAGAGCCCGGAGCAGACCAAGCCCACTGAGGACTTG 62
Qy 221 ArgSerTrpGluylsGlnArgAlaProleuArgPheGlnGlyAspSerArgLeuGlu 240
Db 63 CGGAGCTGGGAGAAAGACAGCGAGCCCGCTGAGTTCCAGGGTGAACGCCCGCTGAG 122
Qy 241 GlnSerGlyCySerTrhshisGysValAspGluAsnileGluArgaGaaNhisTryleu 260
Db 123 CAGCTGGCTGCTACCACTTGCGTAGTAGAACAATCGAGAGAAACCACTACTTA 182
Qy 261 AspleuAgllyllegluAenTrhSerGlnPheGlyProGlySerProSerValAla 280
Db 183 GATTCGCCCGGATGAAATACACATCCCAATTTGGCTGAGCTCCCTCCGTCGCGC 242
Qy 281 GlnlySerGlnleuProProArGTrhSerAsnProThraArgSerArgSerHisGluPro 300
Db 243 CAGAGTCAAGACTGCCCCCGCAGCTCCCACTCGATCTCGCTCCCATGAGCCG 302
Qy 301 GlnAlaIlehisIleProhisArgGlyPProGlnGlyValAspProAlaSerPhehisPhe 320
Db 303 GAAGCCATCCACATCCACCCAGAACCCCAAGCGGTGAGCCGCGCTCCCTCCACTTC 362
Qy 321 LeuAspThrProIleAlaIysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
Db 363 CTTGACACCCCAATCGCAGAGTTCAGAGCTCCAGCAACGGCTCCGGGCAACCAAGAC 422
Qy 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db 423 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCAAGGCAAGAGTGTGGGTGGGCAC 482
Qy 361 ValAlaArgGlyAlaArgAsnLysPProleuGlyProAlaIleProAlaValaSerPro 380
Db 483 GTGGCCAGAGGGGCAAAACMACGCCCTCTGGAGCCCGCAATCCCTGCGTGTCCCC 542
Qy 381 SerAlaHisLeuAlaAspSerProAlaLeuLeuProSerLeuAlaProleuGlyHisLys 400
Db 543 TCCGCCCACTGGGTGCGCAGCGCCGCTCTCCCTCCCTAGGCCCCCTCGGGCAACAG 602
Qy 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProleu 420
Db 603 AAGCACAAGCAACGAGCCAGAGAGCAAGCGGCTGCGGGGCTGCAAGGCAACCACTG 662
Qy 421 AlaSerGlyGlyProValLeuGlyValArgGluHisAlaLeuArgGlyLeuProAlaLeuVal 440
Db 663 GCGCTCAGGTGGCCCTGTCTGGGGGCGGAGCACTCGGGAGTCCCGCTTGTGGTGG 722
Qy 441 TyrglnSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisGlnHis 460
Db 723 TATGAGAGCCAGGCGGGGCAACCGGTCCAGAGCAATGAGCACCAACCACTGAACAT 782
Qy 461 HisHisHisTrhshisHisPheTrhGlnThr 470
Db 783 CACCACCATTAACCACTTCTACAGACA 812

```

RESULT 7

B0645656

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B0645656 921 bp mRNA linear EST 15-JUN-2002
 AGENCOURT_8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
 5', mRNA sequence.
 B0645656 GI:21769828
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 921)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: HICM2484 row: d column: 05
 High quality sequence start: 16
 High quality sequence stop: 685.
 Location/Qualifiers
 1..921
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6285268"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 100"
 /note="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into BscRI/XhoI sites using the following 5' adaptor:
 GGCAGCAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH-MGC
 Library."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:	4.13e-106	Length:	921
Score:	1436.00	Matches:	266
Percent Similarity:	98.89%	Conservative:	1
Best Local Similarity:	98.52%	Mismatches:	3
Query Match:	57.46%	Indels:	0
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x B0645656 (1-921)

```

Qy 201 ASnglnAlaAspleuGInserAlaArgProArGAlaGluThrlyPProThrgluAspleu 220
Db 22 CATCAGGCTGACTGCGAGAGCGCAGAGCCCGGAGCAGACCAAGCCCACTGAGGACTTG 81
Qy 221 ArgSerTrpGluylsGlnArgAlaProleuArgPheGlnGlyAspSerArgLeuGlu 240
Db 82 CGGAGCTGGGAGAAAGACAGCGAGCCCGCTCAGGTTCCAGGGTGAACGCCCGCTGAG 141
Qy 241 GlnSerGlyCySerTrhshisGysValAspGluAsnileGluArgaGaaNhisTryleu 260
Db 142 CAGCTGGCTGCTACCACTTGCGTAGTAGAACAATCGAGAGAAACCACTACTTA 201
Qy 261 AspleuAgllyllegluAenTrhSerGlnPheGlyProGlySerProSerValAla 280

```

```
Db      202 GATCTCGCGGGATGAGAACTACAGCTCCCAATTGGGCTCGCTCCCTCGGTGCC 261
Qy      281 GlnlySerGluLeuProProArgThrSerAnProThraSerArgSerHisGluPro 300
Db      262 CAGAAAGTCAAGACTGCCCCCGGACCTCCCAATCCCACTCGATCTCGCTCCCATGAGCCG 321
Qy      301 GluAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320
Db      322 GAAGCCATCCACATCCCAACCGAAAGCCCAAGCCGTGAGCCGCGCTTCCTTCACCTTC 381
Qy      321 LeuAspThrProIleAlaValSerGluLeuGlnGlnArgLeuAspGlyThrGlnAsp 340
Db      382 CTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGCAACGAGAC 441
Qy      341 GlySerIleHisIlePheValArgSerProGlyAlaGlnGlySerValGlyValGlyHis 360
Db      442 GGGAGCAGACCTTTGTGAGGTCCCCCAAGGCCCAAGGAGTGTGGGTGGCCAC 501
Qy      361 ValAlaArgGlyAlaArgAnlyProProLeuGlyProAlaIleProAlaValSerPro 380
Db      502 GTGGCCAGAGGGGCAAGAAACAAAGCCCCCTGTGGAGCCCGCATCTCGCGGTGCCCC 561
Qy      381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
Db      562 TCGGCCACCTGGCTGCGACGCCGCCCTCTCTCCCTAGCCCCCTCGGCAACAG 621
Qy      401 LysHisIleValArgAlaValGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      622 AACACACAGACGAGCGACAGAGAGCAGACAGGAGCTGCGGGGCTGACAGGACCACTG 681
Qy      421 AlaSerGlyGlyProValLeuGlyArgGluHisIleuArgGlyLeuProAlaValVal 440
Db      682 GCTCTAGGTGCTCTGTCTGTGGGGCGGAGACCTGTGGAGAGCTGCGCGCTGTGGTGTG 741
Qy      441 TyrGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHisIleHisIleHisIleHis 460
Db      742 TATGAGAGCCAGGCGCGGAGCGCGGTCCACAGACATAGACACCAACCATGAAACAT 801
Qy      461 HisHisIleTyrHisIleHisIlePheTyrGlnThr 470
Db      802 CACCACCATTAACCAACCATCTACACAGACA 831

RESULT 8
LOCUS   B0064678                      990 bp      mRNA      linear      EST 02-Apr-2002
DEFINITION AGENCOURT 6853565 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5926427
ACCESSION B0064678
VERSION   B0064678.1 GI:19893537
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 990)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLES    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Struhsberg, Ph.D.
           Email: cgaabs-remail.nih.gov
           Tissue Procurement: Lou Staudt
           cDNA Library Preparation: Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at:
           http://image.llnl.gov
           Plate: LNCM209 row: 1 column: 12
           High quality sequence stop: 670.
FEATURES
     source
         1..990
         /organism="Homo sapiens"
         /mol_type="mRNA"
```

```

/db_xref="taxon:9606"
/clone="IMAGE:5926427"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH MGC 99"
/notes="Organ: Lymph; Vector: pORF7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

ORIGIN
Alignment Scores:
Pred. No.:      9,08e-105      length:      990
Score:          1420.00      Matches:      268
Percent Similarity: 98.53%      Conservative: 0
Best Local Similarity: 98.53%      Mismatches: 2
Query Match:    56.82%      Indels:      2
DB:             Gaps:        0

US-09-993-966-7 (1-470) x B0064678 (1-990)

Qy      201 AasnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrgluAspLeu 220
Db      3 AATCAGGCTGACTGCGAGCGGAGCGGAGGCCCGGAGCAGAGACCAAGCCCATGAGAGACTG 62
Qy      221 ArgSerTrpGlnIlySerGlnAlaGlnAlaProLeuArgPheGlnIlyAspSerArgLeuGlu 240
Db      63 CGAGAGCTGGAGAGAAAGACGACGAGCCCGCTCAGGTTCCAGGGGTGACACCCGCTGGAG 122
Qy      241 GlnSerGlyCysTyrHisIleCysValAspGluAsnIleGluArgAspAsnHisTyrLeu 260
Db      123 CAGTGTGGCTGCTACCACTTCGCTGATGAAACATCCAGAGAGAAACCACTACTTA 182
Qy      261 AspLeuAlaGlyIleGluAsnTyrTrpSerGlnPheGlyProGlySerProSerValAla 280
Db      183 GATCTGCGCGGAGTAAAGAACTACAGCTCCCAATTTGGGCTCGCTCCCTCGGTGCC 242
Qy      281 GlnlySerGluLeuProProArgThrSerAnProThraSerArgSerHisGluPro 300
Db      243 CAGAAAGTCAAGACTGCCCCCGGACCTCCCAATCCCACTCGATCTCGCTCCCATGAGCCG 302
Qy      301 GluAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320
Db      303 GAAGCCATCCCATCCCAACCGAAAGCCCAAGGGGTGAGCCGGGCTCTTCACCTTC 362
Qy      321 LeuAspThrProIleAlaValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
Db      363 CTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGAAAGCGTCCGGGCAACCGAGAC 422
Qy      341 GlySerIleHisIlePheValArgSerProGlyAlaGlnGlySerValGlyValGlyHis 360
Db      423 GGGAGCAGACCTTTGTGAGGTCCCCCAAGGCCCAAGGAGTGTGGGTGGGCGCAC 482
Qy      361 ValAlaArgGlyAlaArgAnlyProProLeuGlyProAlaIleProAlaValSerPro 380
Db      483 GTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAGCCGCGCATCTCCGCGGTGCCCC 542
Qy      381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
Db      543 TCGGCCACCTGGCTGCGACGCCGCCCTCTCTCCCTTACGCCCCCTCGGAGCAAG 602
Qy      401 LysHisIleValArgAlaValGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      603 AACACACAGACGAGCGACAGAGAGCAGACAGGAGCTGCGGGGCTTGACAGGACCACTG 662
Qy      421 AlaSerGlyGlyProValLeuGlyArgGluHisIleuArgGlyLeuProAlaValVal 440
Db      663 GCTTCAAGTGGCCCTGTCTGTGGGCGGAGACACTCTCGGAGAGCTGCCCGGCTTGTGGTGTG 722
```


Qy 441 TyrGluSerGlnAla-GlyGlnProValGlnArgHisGluHisHisHisGluHis 460
 Db 723 TATGAGGCCAGCGCGGAGCGCGGTCCAGAGCATGAGACCAACCACTGATGACA 782

Qy 460 HisHisHisTyrHis-HisPheTyrGlnThr 470
 Db 783 TCACACCATTTACACCACTTCTACCAAGACA 814

RESULT 9
 B0644360
 LOCUS
 DEFINITION AGENCOURT_8351368 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286917
 5', mRNA sequence.
 B0644360
 VERSION B0644360.1 GI:21768532
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 895)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLCM2488 row: h column: 22
 High quality sequence stop: 622.
 Location/Qualifiers
 1. 895
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6286917"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_100"
 /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCGACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

ORIGIN
 Alignment Scores:
 Pred. No.: 5.76e-103 Length: 895
 Score: 1397.00 Matches: 261
 Percent Similarity: 98.13% Conservative: 1
 Best Local Similarity: 97.75% Mismatches: 5
 Query Match: 55.90% Indels: 0
 Gaps: 0
 DB: 5
 US-09-993-966-7 (1-470) x B0644360 (1-895)

Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlyProThrGluAspLeu 220
 Db 4 CATGAGCTGACTGCTGCAAGCGCAAGGCCCGAGCAGACCAACCCATGAGAGACTG 63

Qy 221 ArgSerTyrGluIlyGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 Db 64 CGAGCTGGAGGAAGAAGCAGAGCCCGCTCAAGGTTCCAGGGTGAACAGCCGCTGGAG 123

Qy 241 GlnSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
 Db 124 CAGTCTGGCTCTTACCACTGCGTATGATGAACAATCGAGAGAAACCACTACTTA 183

Qy 261 AspLeuAlaGlyIleGluAsnTyrTrpSerGlnPheGlyProGlySerProSerValAla 280
 Db 184 GATCTGCGCGGATGAAAGAACTACAGCTCCCAATTTGGGCTGCTCCCTCCGTGGCC 243

Qy 281 GlnIlySerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
 Db 244 CAGAGTCAGAACACTGCCCCCGCACCTCCAACTCCACTGATCTGCTCCATGAGCCG 303

Qy 301 GlnAlaIleHisIleProHisArgIlySerProGlnGlyValAspProAlaSerPheHisPhe 320
 Db 304 GAAGCATTCACATTCACACCAAGAGCCCAAGGCGTGAACCCGCTCTTCCACTTC 363

Qy 321 LeuAspThrProIleAlaIlyValSerGluLeuGlnIlyArgLeuArgGlyTyrGlnAsp 340
 Db 364 CTGGACACCCCAATCGCCAGAGTCTCAGAGCTCCAGCAACGGCTCCGGGACACCAAGAC 423

Qy 341 GlySerIlyHisPheValArgSerProIlyAlaGlnGlyIlySerValGlyValGlyHis 360
 Db 424 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCCAAGGAGAGTGGGTGTGGGCCAC 483

Qy 361 ValAlaArgGlyAlaArgAsnIlyProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 484 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGAGCCGCCATCTCGGTGTCTCCCC 543

Qy 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
 Db 544 TCGGCCACCTGGGTGCAACCGCGGCTCTCCCTCCCTTACGCCCCCTCGGGCAACG 603

Qy 401 IlyHisIlyHisArgGlyAlaIlyGlnIlyGlnGlnGlyCysArgGlyIlyLeuGlnAlaProLeu 420
 Db 604 AAGGACAAAGCACCGAGCGAGGAGCGACAGCGCTCGGGGCTCGAAGACCACTG 663

Qy 421 AlaSerGlyIlyProValIlyArgGlyIlyArgGluHisIlyLeuArgGluLeuProAlaLeuVal 440
 Db 664 GCCTCAGGTGGCTCTGTCTGGGGCGGAGACCTCGGGAGCTGCCCTTGCTGTG 723

Qy 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisGluHis 460
 Db 724 TATGAGGCCAGCGCGGAGCGGTCCAGAGCATGAGACCAACCACTGAGACAT 783

Qy 461 HisHisHisTyrHisHisPhe 467
 Db 784 CACCACCTTACCACTTT 804

RESULT 10
 B0644956
 LOCUS
 DEFINITION AGENCOURT_8511562 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6297032
 5', mRNA sequence.
 B0644956
 VERSION B0644956.1 GI:21769128
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 928)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.jnl.gov
 Plate: LHCN2505 row: n column: 09
 High quality sequence start: 17
 High quality sequence stop: 551.
 Location/Qualifiers

FEATURES

source

```

1..928
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6297032"
/cisue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/mode="Organ: Liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA Synthesis Kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 9,99e-102 Length: 928
Score: 1382.00 Matches: 261
Percent Similarity: 96.68% Conservative: 1
Best Local Similarity: 96.31% Mismatches: 7
Query Match: 55.30% Indels: 2
DB: 5 Gaps: 0

```

US-09-993-966-7 (1-470) x BQ644956 (1-928)

```

QY 201 AaenGlaAlaApLeuGlnSerAlaArgProArgAlaGluThrIlyProThrGluApLeu 220
DB 18 CATCAGGCTGACTCTGCAGAGCGCAGGCCCGGAGCAGACCAAGCCCATGAGGACTTG 77
QY 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlu 240
DB 78 CGAGAGCTGGGAGAAAGAGCAGCGCCCGCTCAGTTCCAGGGGTGACAGCCGCTGGAG 137
QY 241 GlnSerGlyCysTrpTrpHisGlyCysValAspGluAsnIleGluArgAsnHisTrpLeu 260
DB 138 CACTCTGGCTGCTACCACTTGGTGTAGTGAACATCCAGAGGAAACCACTACTTA 197
QY 261 AspLeuAlaGlyIleGluAsnTrpTrpSerGlnPheGlyProGlySerProSerValAla 280
DB 198 GATCTGCGCGGATGAGAAACTACACGTCACCAATTTGGCCTGCTCCCTCCGTGCC 257
QY 281 GlnIlySerGluLeuProProArgTrpSerAsnProThrArgSerArgSerHisGluPro 300
DB 258 CAGAAGTCAGAACTGCCCCCGGACCTCCAAATCCCATGATTCGCTCCCATGAGCCG 317
QY 301 GluAlaIleHisIleProHisArgIlyProGlnIlyValAspProAlaSerPheHisPhe 320
DB 318 GAAGCATCCACATCCACACCAAGGCCCAAGGCGCTGAGACCGGCTCTCCACTTC 377
QY 321 LeuAspThrProIleAlaIlyValSerGluLeuGlnIlyProLeuArgGlyTrpThrAsp 340
DB 378 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAAGCGCTCCGGGACACCCAGAC 437
QY 341 GlySerIlyHisPheValArgSerProIlyAlaGlnIlyIlySerValIlyIlyHis 360
DB 438 GGGAGCAGAGCACTTTGTGAGGTCCCGCAAGGCCCAAGGCAAGGTGGGTGGGCGCAC 497
QY 361 ValAlaArgGlyAlaArgAsnIlyProProLeuGlyProAlaIleProAlaValSerPro 380
DB 498 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGAGCCCGCCATCCCTGGGTGTCCCCC 557
QY 381 SerAlaHisIleuAlaIleAserProAlaLeuLeuProSerLeuAlaProLeuGlyIlyHis 400
DB 558 TCGGCCCACTGGCTGCGAGCCGCGGNCCTCTCCCTCCCTAGGCCCCCTCGGGACACAG 617

```

```

QY 401 LysHisIlyHisArgAlaIlyGluSerGlnGlnIlyCysArgGlyIleuGlnAlaProLeu 420
DB 618 AAGCAGACAGACCGAGCGAGAGAGAGCAGAGGCTCCGGGCGCTGCAGGACCACTG 677
QY 421 AlaSerGlyIlyProValIleuGlyArgGluHisIleuArgGluLeuProAlaLeu-ValIva 440
DB 678 GCTTCAGGTGACCTGCTGTGGGCGGAGACACTGCGGAGGCTCCGCTGGTGGG 737
QY 440 IlyIlySerGlnAlaGlyIlyInProValIlyGlnArgHisGluHisIlyHis-HisIlyGluH 460
DB 738 GTATGAGAACCCAGCGCGGAGCGGCTCCGAGACATGGGACACACACCATCATTAC 797
QY 460 IHisHisIlyTrpHisIlyPheTrpGln 469
DB 798 ATCCACACCATTCACACCATCTTACCAG 826

```

RESULT 11

AY412100 888 bp DNA linear GSS 16-DEC-2003
 LOCUS
 DEFINITION Mus musculus NKD1 gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY412100
 VERSION AY412100.1 GI:39768065
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 888)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
 Todd,M.A., Tanenbaum,D.M., Civeillo,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,
 Adams,M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)

JOURNAL

PUBMED
 14671302
 2 (bases 1 to 888)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
 Todd,M.A., Tanenbaum,D.M., Civeillo,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,
 Adams,M.D. and Cargill,M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 This sequence was made by sequencing genomic exons and ordering
 them based on alignment.

TITLE

JOURNAL
 COMMENT
 FEATUERS
 source
 1..888
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>888
 /gene="NKD1"
 /locus_tag="NCM4444"

ORIGIN

Alignment Scores:
 Pred. No.: 4.18e-101 Length: 888
 Score: 1374.00 Matches: 259
 Percent Similarity: 91.22% Conservative: 11
 Best Local Similarity: 87.50% Mismatches: 26
 Query Match: 54.98% Indels: 0
 DB: 9 Gaps: 0

US-09-993-966-7 (1-470) x AY412100 (1-888)

```

QY 65 GluLeuValGlyAspValIleuArgAspThrLeuSerGlnGluGluIlyuAspPheArg 84
DB 1 GAACGTGTGGGTACACTTCTNAGAGAGGCTCTCGTGAGAGAGACGAGCGACTTCCC 60
QY 85 LeuGluValAlaLeuProProGluIlyTrpAspGlyLeuGlySerGlyAspGlyIlyValys 104

```

```

Db      61 CTGAGATGCGCCCTGCGCCCTGAGAAATGACAGCCTAGGTAGTGAGATGAGAAAGA 120
Qy      105 MetGluArgValSerGluProCysProGlySerIlySerGluLeuIlypneGluLeu 124
Db      121 ATGAGAGAGATGAGCGAAGCTGCGCAGCGCTCCCAAGAAAGACGTCCAGTTGGAAGACTG 180
Qy      125 GluCysAspValSerMetGluIlyAspSerArgGlnIlyPheThrPheThrLeuTyrAsp 144
Db      181 CAGGTATGATGTCCTGTGAGAGAGACAGCCGCGCAAGAGTGAGACTTTCACCTTATATGAC 240
Qy      145 PheAspAsnAsnGlyIlyValThrArgGluAspIleThrSerLeuLeuH1sThrIleTyr 164
Db      241 TTGAGACAAATGAGCAAGTAGACCCGCGAGAGACATTACAGCTTGCTGCAATACCATCTAT 300
Qy      165 GluValValAspSerSerValAsnHisSerProThrSerSerIlyMetLeuArgValIly 184
Db      301 GAAATGCTTACCTCTCTGTGAACCATTTCCCAATCAAGACAGACATGCGGGTGAG 360
Qy      185 LeuThrValAlaProAspGlySerGlnSerIlyArgSerValLeuValAsnGlnAlaAsp 204
Db      361 CTCACCTGCTCCTGACGGAGCCAGAGTAAAGAGAGCGCTTTTCAACCATACCGAT 420
Qy      205 LeuGlnSerAlaArgProAlaArgAlaGluThrIlySerProThrGluAspLeuArgSerTyrGlu 224
Db      421 CTGAGAGACACAAAGCCCGCAGACAGACACCAACCCGCTGAGAGCTGCGGTGGAG 480
Qy      225 IlyIlyGlnIlyArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlnIlySerGlyCys 244
Db      481 AAGAGACAGCGAGCCCACTCAGGTTCCAGGTTCCAGGTTCCAGGTTCCAGGTTCCAGGTTCC 540
Qy      245 TyrHisGlyCysValAspGluAsnIleGluArgArgAsnH1sTyrIleAspLeuAlaGly 264
Db      541 TACACCATTTGCGGATGAGAAATGAGAGAGAAACCACTACCTAGACCTGCGGAG 600
Qy      265 IleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAlaGlnIlySerGlu 284
Db      601 ATGAGAGACTACGCTCAGCTGAGTTGAGACCGGAGTCCCTTCGTTGCGCCAGAACTCAGAG 660
Qy      285 LeuProProAlaArgThrSerAsnProThrArgSerArgSerH1sGluProGluAlaIleH1s 304
Db      661 CTGCGCCCTCGAATCTCAACCCCACTCGCTCTGCTCCCAAGAGCAAGAGCTGCCAC 720
Qy      305 IleProH1sArgIlySerProGlnIlyValAspProAlaSerPheH1sPheLeuAspThrPro 324
Db      721 ATCCCAACACGAGAGCCCAAGGTGTGAGACCAAGCTCTCTTCCACCTCTTGACACCCCA 780
Qy      325 IleAlaIlyValSerGluLeuGlnIlyArgLeuArgIlyThrGlnAspGlySerIlyH1s 344
Db      781 TTTCGCAAGGATCAGAGCTCCACCAACGCTCCGGGGCACTCAGATGGAGACAGAC 840
Qy      345 PheValIlyArgSerProIlyAlaGlnIlyIlySerValGlyValGlyH1s 360
Db      841 TTGTGTAGTCTCCCAAGGCCCAAGGCAAGAAACATGGGTATGGGCCAC 888

RESULT 12
BQ646371 1008 bp mRNA linear EST 15-JUN-2002
LOCUS BQ646371
DEFINITION AGENCOURT 8493065 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6300188
5', mRNA sequence.
ACCESSION BQ646371
VERSION BQ646371.1 GI:21770543
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1008)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs-remail.nih.gov

```

```

FEATURES
    source
        1..1008
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6300188"
            /tissue_type="hepatocellular carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_1lb="NIH MGC 100"
            /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-cDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH_MGC
            Library."

ORIGIN
Alignment Scores:
Pred. No.: 1,09e-93 Length: 1008
Score: 1283.50 Matches: 254
Percent Similarity: 93.45% Conservative: 3
Best Local Similarity: 92.36% Mismatches: 11
Query Match: 51.36% Indels: 7
DB: Gaps: 2

US-09-993-966-7 (1-470) x BQ646371 (1-1008)
Qy      201 AsnGlnAlaAspLeuGlnSerAlaArgProAlaGluThrIlySerProThrGluAspLeu 220
Db      22 CATCAGGCTGACTGACGACGAGCGAGCCCGCAGAGACCAAGCCCACTGAGAGACTG 81
Qy      221 ArgSerTyrGlnIlySerGlnIlyArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlu 240
Db      82 CGAGAGCTGGAGAAAGAGAGAGCGAGCCCGCTCAGAGTTCCAGGGGTGACGCGCTGAG 141
Qy      241 GlnSerGlyCysTyrHisCysValAspGluAsnIleGluArgArgAsnH1sTyrLeu 260
Db      142 CAGTCTGGCTGCTACCACTTCGTTGATGAGAAACATCGAGAGAAACCACTATTA 201
Qy      261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db      202 GATCTCGCGGAGTAGAAACTACACGCTCCCAATTTGGGCTGCTCCCTTCGTGGCC 261
Qy      281 GlnIlySerGluLeuProProAlaArgThrSerAsnProThrArgSerArgSerH1sGluPro 300
Db      262 CAGAAAGCAAAATGCGCCCGCAGACTCCCAATCTCCATGATCTGCTCCCATGAGACCG 321
Qy      301 GluAlaIleHisIleProH1sArgIlySerProGlnIlyValAspProAlaSerPheH1sPhe 320
Db      322 GAAAGCTTCACATCCCAACCGAAAGCCCAAGGCTGAGACCGGCTCTTCACTTC 381
Qy      321 LeuAspThrProIleAlaIlyValSerGluLeuGlnIlyArgLeuArgIlyThrGlnAsp 340
Db      382 CTGAGACACCCCAATCGCAAGGTCTCAGAGCTCCAGCAACGAGCTCCGGGCAACCAAGAG 441
Qy      341 GlySerIlyHisPheValIlyArgSerProIlyAlaGlnIlyIlySerValGlyValGlyH1s 360
Db      442 GGGAGCAAGCACTTTGTGAGAGTCCCGCAGAGGCCCAAGGCAAGATGTGGGTGGGCCAC 501
Qy      361 ValAlaArgGlyAlaArgAsnIlySerProLeuGlyProAlaIle-ProAlaValSerPr 380

```

```

Db      502 GTGGCCAGAGGGGCAAGAAACACACCCCTCTGGAGCCGACATCCCTGGGTGTCC 561
Qy      380 oSerAlaHleuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHleu 400
Db      562 CTCGGCCACCTGGCTGCGAGCCCGGCGCTCTCCCTCCCTACCCCGGCGACAA 621
Qy      400 gLyHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleu 420
Db      622 GAAGGACACACACGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 681
Qy      420 uAlaSerGlyGlyProVal-LeuGlyArgGlyHleuHleuHleuHleuHleuHleuHleu 438
Db      682 GGCCTCAGGTGGCCCTCTCTCTGGAGGAGACCTTGCGAGCTGCGCGCTGG 741
Qy      439 ---ValValTyrgLuserGlnAlaGlyGlnPro---ValGlnArgHleuHleuHleuHleu 456
Db      742 GGGGGGGGTATGAAGACCGGCGGCGGAGAGCGGGGTCCAGAGACATGGAGCACACAC 801
Qy      457 -HleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleu 468
Db      802 CCACCATGAACATCACCCCATTTACACACTTCTAC 838

RESULT 13
B0652087 996 bp mRNA linear EST 15-JUL-2002
LOCUS     AGENCOURT_8207495 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6283110
DEFINITION
5', mRNA sequence.
ACCESSION B0652087
VERSION    B0652087.1 GI:21776259
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 996)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            plate: iImage2478 row: j column: 07
            High quality sequence stop: 604.
FEATURES
    source
        1..996
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6283110"
            /tissue_type="hepatocellular carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_1lb="NIH MGC 100"
            /note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
            EcoRI; CDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-CDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH-MGC
            Library."
ORIGIN
Alignment Scores:
Pred. No.:      3,85e-91      Length:      996
Score:          1252.00      Matches:      243

```

```

Percent Similarity: 94.27%      Conservative: 4
Best Local Similarity: 92.75%      Mismatches: 12
Query Match: 50.10%      Indels: 3
DB: 5      Gaps: 2

US-09-993-966-7 (1-470) x B0652087 (1-996)

Qy      201 AAGGlnAlaApleuGlnSerAlaArgProAlaGlyAlaThrLyProThrGlyApleu 220
Db      3 AATCCGCTGACTGTCAGACCGGAGAGCCCGGAGACAGACCAAGCCACTGAGACTG 62
Qy      221 ArgSerTrpGlyLysLysGlnArgAlaProLeuArgPheGlnGlyAAsSerArgLeuGly 240
Db      63 CGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 122
Qy      241 GlnSerGlyGlyTrpHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleu 260
Db      123 CAGTCTGGCTGTACACCACTTGGTATGAGAACATCGAGAGAGAGAGAGAGAGAGAGAG 182
Qy      261 ApleuAlaGlyLleGlyAAsTrpHleuHleuHleuHleuHleuHleuHleuHleuHleu 280
Db      183 GATCTCGCGGAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
Qy      281 GlnLysSerGlyLysProArgTrpSerArgProThrArgSerArgSerHleuPro 300
Db      243 CAGAGTCCAGAGACTGCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
Qy      301 GlnAlaHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleu 320
Db      303 GAAGCATTCACATCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
Qy      321 LeuAspThrProIleAlaLysValSerGlyLeuGlnArgLeuArgGlyThrGlnAsp 340
Db      363 CTTGACACCCCAATCCGCAAGGTCTCAGAGCTCAGACAGAGAGAGAGAGAGAGAGAGAG 422
Qy      341 GlnSerTrpHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleu 360
Db      423 GAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
Qy      361 ValAlaArgGlyAlaArgAAsLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db      483 GTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542
Qy      381 SerAlaHleuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHleu 400
Db      543 TCCGCGCCACTGGCTGTCAGAGCCCGGCGCTCTCCCTCCCTACAGAGAGAGAGAGAG 602
Qy      401 LysHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleuHleu 420
Db      603 AAGCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
Qy      421 AlaSerGlyGlyProValLeuGlyArgGlnHleuHleuHleuHleuHleuHleuHleuHleu 439
Db      663 GCTCAGGTGGCCCTGTCTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
Qy      440 ValTyrgLuserGlnAlaGlyGlnProVal---GlnArgHleuHleuHleuHleuHleu 458
Db      723 GATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
Qy      458 gGly 459
Db      783 TGA 786

RESULT 14
LOCUS     AY412099 885 bp DNA linear GSS 16-DEC-2003
DEFINITION Pan troglodytes NKD1 gene, VIRTUAL TRANSCRIPT, partial sequence,
ACCESSION AY412099
VERSION    AY412099.1 GI:39768064
KEYWORDS   GSS.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes

```

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan. 1 (bases 1 to 885) Clark,A.G., Gienowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeallo,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.C., Adams,M.D. and Cargill,M.
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL	Science 302 (5652), 1960-1963 (2003)
PUBMED	14671302
REFERENCES	2 (bases 1 to 885) Clark,A.G., Gienowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeallo,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.C., Adams,M.D. and Cargill,M.
TITLE	Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
JOURNAL	This sequence was made by sequencing genomic exons and ordering them based on alignment.
COMMENT	Location/Qualifiers
FEATURES	source 1..885 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1...>885 /gene="NKD1" /locus_tag="HCN4444"
ORIGIN	
Alignment Scores:	
Pred. No.:	3,06e-90 Length: 885
Score:	1240.00 Matches: 247
Percent Similarity:	84.41% Conservative: 2
Best Local Similarity:	83.73% Mismatches: 46
Query Match:	49.62% Indels: 0
DB:	Gaps: 9
US-09-993-966-7 (1-470) x AY412099 (1-885)	
Oy	65 G U E U E A G Y A P V A L E U A G A P T R L E U S E G I U G I U A S P H E A T G 84
Db	1 GAGCTCGTGCGCCACGCTGTGAAGAACAAGCTCAGCGAGCAAGAAAGAACAGACGACTTTCCG 60
Oy	85 L E U G I U A L A E U P P O G I U Y T R A P G I Y L E U G I S E G I A S P G I U E Y S 104
Db	61 CTGGAAATGGCCCTTCGCTCTGTGAAGACTGACGGGCTGGAGCGCGAGATGAGAGAG 120
Oy	105 M E G I U A R Y A S E G I U P P O Y C P P O J Y S E T Y S G I N L E U Y P H E G I U E U 124
Db	121 A T G A G A G A G T G A G C A C C T G C C A G G C T C A A G A G C A G C T T T G A A G A G C T 180
Oy	125 G I N C Y A S P A S P A S E T E T G I U G I U A P S E A T G I U T T P T R P H E T R L E U Y T A S P 144
Db	181 C A G G C A C G T G C G T G G A G G A C A C C G C G A G T G G A C C T T C A C C T G A T G A C 240
Oy	145 P h e a s p h e n g i l y s v a t h r a x g u a p i e t h r s e r l e u h i:e t h r l c t r 164
Db	241 T T T A C A C A A C G C A A G G T C A C C C G A G A C A T C A C G T T G T C T C A C A C A T C A T 300
Oy	165 G I U A I V A A S P S E S E R V A A h h i s e r P r h t S e r S e r Y M e t L e u A r Y A l l ys 184
Db	301 G A G T G T G A C T C T G T C A C C A C C T C C C A A C A T C C A G A A A T G C T G G G T A A G 360
Oy	185 L e u t h y A A A P P o A P G I Y S E G I U S E r Y A T G S E r V A L L e u A A h A h A A S P 204
Db	361 C T A C C C T G G C C C C G A T G C A C C G A C A A G A G A G G C T G T G T C A A T C A G G C T G A C 420
Oy	205 L e u G I S E r A A A P P o A P G I A G I U T R Y S P P r h t G I U A S P L e u A A S E r T T G I U 224
Db	421 C T G C A G G C C A A G G C C C G A G A G A C C A A G C C C A C T G A G A C T T C G A G C T G G G A G 480

OY		225	Lysylserglnarginalaproleualrgphaglnglyaspsearlgleuglinserglycys	244
Dd		481	AAGAAGCAGCGAGGCCGGCTCCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	540
OY		245	Tyrhishtcyvalaapgluaenllgeluargrghasnhsttrytleuaspleualaigly	264
Dd		541	NN	600
OY		265	IlegluabntyrtThrserGlInPhegLyProGlySerProSeraValaglntylerSergLu	284
Dd		601	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCCCCCTTCGCggccCAgaactCAGAA	660
OY		285	LeudroProlArqThrSrAsanPrOfChrArgSerArSgerHtsicguLproGuialatIHis	304
Dd		661	CtgcCCCCCcgcAcctCCAATCCACACTCGATTGTGCATCATGAAGCCGaAACCAtCAC	720
OY		305	IlEprhoIsaRgyvsProInglyVAlAspProlaSeRpheHlsPheleuAsPTHePro	324
Dd		721	GTTCCACACCGAAGAACCCCAGAggtGacCCGGScTCTTCACttCTTGACAccccca	780
OY		325	IleaIalysVaISergiuleGngLnArgleuarSglyThrGlnaeRpGIyerLYeShis	344
Dd		781	ATCCCAAAGTCtCAGAGCTCCAGCAACGGCTCCGGSGCACCCAGAGCGGAGCAACAC	840
OY		345	PheVALARgsER-POLYSALagInGLYLyseRVAlaly 359	
Dd		841	TTTTGTGAGGTCCCCCAAGGCCAGGGGNnnnNTGTGTGTGTGGAGC	885
RESULT_15		B0649813	966 bp mRNA linear EST-15-JUN--2002	
LOCUS		B0649813		
DEFINITION		AGENTCOURT_830249 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271026		
ACCESSION		B0649813		
VERSION		B0649813.1		
KEYWORDS		EST.		
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Caniata; Vertebrata; Euarchontomi;			
REFERENCE	Eutharieta; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	1 (bases 1 to 966) NIH-MGC http://mgc.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Straubeberg, Ph.D. Email: Ggapds-iemail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMN at: http://image.lnl.gov plate: LCM2447 row: b column: 19 High quality sequence stop: 583.			
FEATURES	location/Qualifiers			
SOURCE	1..966			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6271026"			
	/tissue_type="hepatocellular carcinoma, cell line"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_idb="NIH_MGC_100"			
	/note="Organ: liver; Vector: pOTB7, Site_1.XhoI, Site_2:EcoRI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor:			
	GGAACGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library."			

ORIGIN

Alignment Scores:

Pred. No.:	3,066-88	Length:	966
Score:	1216.00	Matches:	233
Percent Similarity:	97.91%	Conservative:	1
Best Local Similarity:	97.49%	Mismatches:	4
Query Match:	48.66%	Indels:	1
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x BQ649813 (1-966)

```
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
DB 2 CATGAGCTGACCTTGCAGAGCGCAAGGCCCGAGCAGACCAAGCCACCTGAGGACCTG 61
QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
DB 62 CGGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
QY 241 GlnSerGlyCysTrpLysHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
DB 122 CAGTCTGGCTGCTACCACTTCGATGATGAGAAACATCGAGAGAGAGAAACACTACTTA 181
QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
DB 182 GATCTCGCCGGGATGAAACTACACGTCCTCAATTTGGGCTGGCTCTCCCTTCGAGGCC 241
QY 281 GlnLysSerGluLeuProArgTrpSerAsnProThrArgSerArgSerHisGluPro 300
DB 242 CAGAGTCAGAACTGCCCCCGGCACTCCATCCCACTGATCTCGCTCCCATGAGCCG 301
QY 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
DB 302 GAAGCCATCCACATCCACACCGAAAGCCCAAGGCGTGAGCCGCGCTCTTCCACTTC 361
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLysArgGlyThrGlnAsp 340
DB 362 CTGACACCCCAATGCCAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGGCAACCAGGAC 421
QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
DB 422 GGGAGCAAGCACTTGTGAGGTCTCCCAAGGCCCAAGGCAAGAGTGTGGTGGCCAC 481
QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
DB 482 GTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGGAGCCCGCATCCTGCGGTGTCCCC 541
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
DB 542 TCCGCCCACTGGCTGCAAGCCCGGGCTCTCTCCCTCCCTAGCCCCCTCGGGCACAG 601
QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
DB 602 AAGCACAAGACCAAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
QY 421 AlaSerGlyGlyProValLeuGlyArgGluHisLeu-ArgGluLeuProAlaLeu 438
DB 662 GCCTCAGGTGGCTGTCTGTGGGGGAGAGACACTGCGGAGGTTGGCGGCTTG 716
```

Search completed: December 30, 2004, 07:54:26
Job time : 5660 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 03:48:50 ; Search time 5778 Seconds

(without alignments)
459.446 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 2499
Sequence: 1 MGLKSKPAVCKGRSPG.....RHHHHHHHHHHHHFYQT 470

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4176236 seqs, 2824127955 residues

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-O/cgn2_1/USPTO.spool.p/US0993966/runat.29122004.130759.9062/app.query.fasta_1.647
-DB=Published Applications NA -QFMT=fasta -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -UNIT=bits -START=1 -THR SCORE=pct -THR MAX=100
-TRANS=human40.ccl -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0993966@cgn2_1.1.912 @runat.29122004.130759.9062
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US09_PUB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2499	100.0	1448	13 US-10-087-192-179	Sequence 179, App
2	2499	100.0	1786	10 US-09-993-966-1	Sequence 1, Appl1
3	2499	100.0	1859	10 US-09-993-966-5	Sequence 5, Appl1
4	2189.5	87.6	1416	10 US-09-993-966-2	Sequence 2, Appl1
5	2171.5	86.9	1401	10 US-09-993-966-6	Sequence 6, Appl1
6	2167.5	86.7	1401	9 US-09-730-989-1	Sequence 1, Appl1
7	2085.5	83.5	2807	13 US-10-087-192-176	Sequence 176, App
8	1191.5	47.7	106315	13 US-10-087-192-178	Sequence 178, App
9	1076	43.1	98474	13 US-10-087-192-175	Sequence 175, App
10	1067	42.7	2379	10 US-09-764-891-5904	Sequence 5902, App
11	1067	42.7	2379	10 US-09-764-891-5904	Sequence 5904, App
12	1053	42.1	590	10 US-09-993-966-21	Sequence 21, Appl1
13	831.5	33.3	1966	9 US-09-764-868-152	Sequence 152, Appl1
14	786	31.5	1959	18 US-10-723-860-5220	Sequence 5220, App
15	739	29.6	427	10 US-09-918-995-8686	Sequence 8686, App
16	378.5	15.1	1817	9 US-09-822-849A-83	Sequence 83, Appl1
17	274	11.0	73038	17 US-10-322-696-37	Sequence 37, Appl1
18	271	10.8	639	15 US-10-243-552-231	Sequence 31, App
19	259	10.4	901	13 US-10-027-632-162722	Sequence 162722, Sequence 162722,
20	259	10.4	901	15 US-10-027-632-162722	Sequence 126, App
21	246	9.8	1950	13 US-10-016-157A-126	Sequence 7990, App
22	239	9.6	598	15 US-10-029-386-7990	Sequence 18, Appl1
23	238	9.5	148	10 US-09-993-966-18	Sequence 20, Appl1
24	237	9.5	128	10 US-09-993-966-20	Sequence 21690, A
25	233	9.3	133	15 US-10-029-386-21690	Sequence 14, Appl1
26	233	9.3	134	10 US-09-993-966-14	Sequence 16, Appl1
27	186	7.4	107	10 US-09-993-966-16	Sequence 17, Appl1
28	175	7.0	96	10 US-09-993-966-17	Sequence 17, Appl1
29	172	6.9	6749	18 US-10-723-860-6868	Sequence 6868, App
30	161.5	6.5	554	15 US-10-029-386-7615	Sequence 7615, App
31	161	6.4	2481	18 US-10-425-115-114040	Sequence 114040,
32	156.5	6.3	1633	15 US-10-017-161-1819	Sequence 1819, App
33	156.5	6.3	1633	15 US-10-029-798-1475	Sequence 1475, App
34	156.5	6.3	15574	13 US-10-026-188-6	Sequence 6, Appl1
35	155	6.2	84428	16 US-10-229-1488-1	Sequence 1, Appl1
36	153	6.1	744802	9 US-10-892-798-1369	Sequence 1369, App
37	152	6.1	2433	5 US-09-944-807-3	Sequence 3, Appl1
38	152	6.1	2433	15 US-10-290-198-2	Sequence 2, Appl1
39	152	6.1	2433	15 US-10-172-118-1068	Sequence 1068, App
40	152	6.1	2433	16 US-10-342-887-1068	Sequence 1068, App
41	152	6.1	2433	18 US-10-874-015-3	Sequence 3, Appl1
42	151.5	6.1	2078	15 US-10-307-047-35	Sequence 35, Appl1
43	151	6.0	3369	18 US-10-786-720-8	Sequence 8, Appl1
44	151	6.0	3477	18 US-10-786-720-7	Sequence 7, Appl1
45	150.5	6.0	2978	14 US-10-116-802-184	Sequence 184, App

ALIGNMENTS

RESULT 1
US-10-087-192-179
Sequence 179, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OR INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 179
LENGTH: 1448
TYPE: DNA

ORGANISM: Homo sapiens
US-10-087-192-179

Alignment Scores:

Pred. No.:	3.52e-227	Length:	1448
Score:	2499.00	Matches:	470
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:	13	Indels:	0
		Gaps:	0

US-09-993-966-7 (1-470) x US-10-087-192-179 (1-1448)

```

QY      1 MetGlylyseuhserrlyspProAlaAlaValCylysaRgaRgluserProglugly 20
Db      39 ATGGGGAACCTTCACTCCAGCCGCGCTGTGCAAGCGCAGAGAGCCGGAAGT 98
QY      21 AspSerPheAlaValSerAlaAlaTPrAlaRglYglYllegluglYTrpIlleglYArg 40
Db      99 GACAGCTTCCCGCTGACGCGCTGCTGGAGGCGATCGAGAGTGGATCGGAGA 158
QY      41 GluArgCyseProglYglYValSerGlyProRglInleuRglueuAlaglYThrIlleglY 60
Db      159 CAGGCGTCCCGGCGGCTGTCTCGGAGCCCGACAGCTGCGGTTGCGGGCACCATTAGGC 218
QY      61 ArgSerThrArgGluLeuValAlglYAspValLeuArgAspThrLeuSerGluIngluGlu 80
Db      219 CGAAGCACCAGGAGCTGTGGGCGAGCTGTGAGACACAGCTCAGCGAGAGAGAG 278
QY      81 AspAspPheArgLeuGluValAlaLeuProProgluYThrAspGlyLeuGlySerGly 100
Db      279 GACGACTTGGCGCTGGAAGTGGCCCTCTCTAGAGAGCTGACGGCTGGGAGCGGA 338
QY      101 AspGluYlyseuhserrlyspProAlaAlaValCylysaRgaRgluserProglugly 120
Db      339 GATGAGAAGAGATGAGAGAGTGAAGCACTCGCCAGGCTCCAGAGAGCACTGAG 398
QY      121 PheGluInleuGlnCyseAspValSerMetGluInleuAspSerAlaGlnYTrpThrPhe 140
Db      399 TTGGAAAGCTCCAGCTCGAGCTGTCCATGAGAGAGACACCGCGAGAGTGAACCTTC 458
QY      141 ThrLeuYlyseuhserrlyspProAlaAlaValSerGlyProRglInleuRglueuAlagl 160
Db      459 ACCCTGATGACTTGTGACACACAGCGCAAGTCAACCGAGAGGACATCACAGCTTGTG 518
QY      161 HisThrIleYrgluValAlaAspSerSerValAsnHisSerProThrSerSerlyMet 180
Db      519 CACACCATATGAGGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAAGT 578
QY      181 LeuArgValYlyseuhserrlyspProAlaAlaValSerGlyProRglInleuRglueuAlagl 200
Db      579 CTGGGGTAAAGCTCACCGTGGCCCGATGGCCAGACAGAGAGAGAGGCTCTTGTG 638
QY      201 AsnGlnAlaAspLeuGlnSerAlaArgProRglInleuRglueuAlaglYTrpThrPhe 220
Db      639 AATAGGCTGACTGCGAGCGCAAGGCCCGAGACAGAGACCACTGAGAGACTG 698
QY      221 ArgSerTrpGluYlyseuhserrlyspProAlaAlaValSerGlyProRglInleuRglueuAlagl 240
Db      699 CGGAGCTGGGAGAGAGAGCGAGCGCCGCTCAAGTTCCAGGCTGACAGCGCTGGAG 758
QY      241 GluSerGlyCylysiYHHisCyseValAspGluAsnIllegluArgAsnHisYTrpLeu 260
Db      759 CAGCTGTGCTGTCCACCATTTGCTAGTAGAACAATCGAGAGAGAAACCACTACTTA 818
QY      261 AspLeuAlaGlyllegluAsnYTrpSerGlnPheGlyProglYserProSerValAla 280
Db      819 GATTCCTCCGGGATGAAATTAACGCTCCCAATTTGGGCTGCTCCCTCCCTGAGCC 878
QY      281 GluYseuhserrlyspProAlaAlaValCylysaRgaRgluserProglugly 300
Db      879 CAGAGCTCAGAACTGCCCGCCGCACTCCATCCATCCGATTCCTCCAGAGCCG 938

```

```

QY      301 GluAlaIleHisIleProHisArgYseProGlnGlyValAspProAlaSerPheHisPhe 320
Db      939 GAGCCATCCACATCCACACCGAAGCCCCAAGGCTGAGACCGGCTCTTCACTTC 998
QY      321 LeuAspThrProIleAlaYlyseuhserrlyspProAlaAlaValSerGlyProRglInleuRglueuAlagl 340
Db      999 CTTGACACCCCAATCCGCAAGGTCTCAGACTCCAGACAGCGCTCCGGGGCACCAGAC 1058
QY      341 GlySerIlysiPheValAspSerProYlyAlaGlnYlyseuhserrlyspProAlaAlaValSerGly 360
Db      1059 GGGAGCAAGCACTTGTGAGGTCCCGCAAGGCCCAAGGCAAGAGTGTGGTGGCCAC 1118
QY      361 ValAlaArgGlyAlaAspGlnYseuhserrlyspProGluGlyProAlaIleProAlaValSerPro 380
Db      1119 GTGGCCAGAGGGGCAAAACAGCCCCCTCTGGAGCCCGCATCTTGGGTCTCCCC 1178
QY      381 SerAlaHisLeuAlaAspProAlaLeuProSerLeuAlaProLeuAlaYlyseuhserrlyspPro 400
Db      1179 TCGGCCACCTGGGTGCAAGCCCGGCTCTCTCCCTCCCTAGCCCGCTCGGACAG 1238
QY      401 LysHisIlysiPheValYlyseuhserrlyspProGlnGlyCyseArgGlyLeuGlnAlaProLeu 420
Db      1239 AAGCACAAGCAGCGAGGCAAGAGAGCAGAGGCTGCCGGGCTGCAAGGCACTG 1298
QY      421 AlaSerGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuValAla 440
Db      1299 GCTCAGGTGACCTGTCTGGGGGAGGACCTGCGGAGCTGCCGCTGTGTGTG 1358
QY      441 TyrGluSerGlnAlaGlyProValGlnArgHisGlnHisIlysiPheGlnHis 460
Db      1359 TATGAGCCAGGCGCGGCGAGCGGTCCAGAGCATATGAGCACCACCATGAAAT 1418
QY      461 HisHisIlysiPheHisPheYrglnThr 470
Db      1419 CACCACCATTAACCACTTCAACAGACA 1448

```

RESULT 2

```

US-09-993-966-1
: Sequence 1, Application US/09993966
: Publication No. US20030186232A1
: GENERAL INFORMATION:
: APPLICANT: ROHMAN, MICHAEL
: TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
: FILE REFERENCE: 014024/0280733
: CURRENT APPLICATION NUMBER: US/09/993,966
: PRIOR FILING DATE: 2001-11-27
: PRIOR APPLICATION NUMBER: 60/252,884
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/291,109
: PRIOR FILING DATE: 2001-05-16
: PRIOR APPLICATION NUMBER: 60/325,571
: PRIOR FILING DATE: 2001-10-01
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 1
: LENGTH: 1786
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-993-966-1

```

Alignment Scores:

Pred. No.:	4.62e-227	Length:	1786
Score:	2499.00	Matches:	470
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-993-966-7 (1-470) x US-09-993-966-1 (1-1786)

```

QY      1 MetGlylyseuhserrlyspProAlaAlaValCylysaRgaRgluserProglugly 20

```



```

Db      159 ATGGGAACTTCACTCCAAAGCCGCGCTGTGCAAGCGAGAGAGAGCCGGAAGCT 218
QY      21  AapSerPheAlaValSerAlaAlaATPAlaArgIlySgIyllegIuIuTPIllegIyArg 40
Db      219 GACAGCTTGCGCTGAGAGCGCTGCTGGGCTCGGAAGGCAATCGAGAGTGGATCGGAGA 278
QY      41  GlnArgCySerProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db      279 CACGCGTGGCGCGGCGGTGTCTGGGACCCCGACAGCTGGCGTGGCGGCAACATAGGC 338
QY      61  ArgSerThrArgGlyLeuValGlyAapValLeuArgAphThrLeuSerGlnGluGlu 80
Db      339 CGAAGACCCCGGAGCTCGTGGCGAGCTGTGAGACACAGCTCAGCGAGAGAGAG 398
QY      81  AapAapPheArgLeuGluValAlaLeuProProGlyIleThrAapGlyLeuGlySerGly 100
Db      399 GACGACTTTCGCTGGAGAGTGGCCCTGCTCTGAGAAAGACTGACGGGCTGGGAGCGGA 458
QY      101 AapGlyLulYlySerMetGluArgValSerGluProCySerProGlySerIlySgIylleg 120
Db      459 GATGAGAAAGAAATGAGAGAGAGAGAGAAACCTGCGCAGAGCTTCAGAAAGCAAGTGA 518
QY      121 PheGlyGlyLeuGlnCySerAapValSerMetGlyGluAapSerArgGlnGluTPIThrPhe 140
Db      519 TTTGAAAGAGCTCAAGTCCGACGCTGTCCATGAGAGAGACCGCGAGAGTGAACCTTC 578
QY      141 ThrLeuTyraPheAapAenAangIlyValThrArgGluAapIleThrSerLeuLeu 160
Db      579 ACCCTGATGACTTGTGACAAACAGCGAGAGTCAACCGAGAGAGCATCACAGCTTCG 638
QY      161 HisThrIleTyrluValAlaAapSerSerValAenHisSerProThrSerSerIlyMet 180
Db      639 CACACCATCATATAGAGGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAATAG 698
QY      181 LeuArgValIlySerThrValAlaProAapGlySerGlnSerIlySgIyllegVal 200
Db      699 CTCGGGTAAAGCTCACCGTGGCCCCCGATGGCGAGCGAGAGAGAGAGAGAGCTTCG 758
QY      201 AangIAlaAapLeuGlnSerAlaArgProArgAlaGluThrIlySgIyllegIuAapLeu 220
Db      759 AATCAGAGCTGACCTGAGAGCGAGAGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGCTG 818
QY      221 ArgSerTrpGlyLulYlySgIyllegIuArgAlaProLeuArgPheGlnGlyAapSerArgLeuGlu 240
Db      819 CGAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
QY      241 GlnSerGlyCySerThrHisGlyValAapGlyAapIleGluArgAapAenHisGlyLeu 260
Db      879 CAGTCTGGCTGCTACCACTTCGCTGATGAGAGACATCGAGAGAGAGAGAGAGAGAGAG 938
QY      261 AapLeuAlaGlyIlegIuAenTyrluSerGlnPheGlyProGlySerProSerValAla 280
Db      939 GATCTCGCGGAGTAAAGAACTACAGCTCCCAATTTGGGCTGGCTCCCTTCGCTGGCC 998
QY      281 GlnIlySerGlyLeuProProArgThrSerAapProThrArgSerArgSerHisGluPro 300
Db      999 CAGAAATTCAGAACTGCCCCCGACCTCCATCTCCATCTCCCTCCATGAGCCG 1058
QY      301 GlnAlaIleHisIleProHisArgIlySgIyllegIuAapProAlaSerPheHisPhe 320
Db      1059 GAGGCACTTCACATCCCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
QY      321 LeuAapThrProIleAlaIlySgIyllegIuGlnGlnArgLeuArgIyluThrGlnAap 340
Db      1119 CTGAGACCCCAATCGCAGAGTCTCAGAGCTCAGAGACGGCTCGGGGACACCAAGAG 1178
QY      341 GlySerIlyHisPheValArgSerProIlyAlaGlnGlySerValGlyValGlyHis 360
Db      1179 GGGAGAGAGAGACTTGTGAGGTCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1238
QY      361 ValAlaArgGlyAlaArgAenIlySgIyllegIuProAlaIleProAlaValSerPro 380
Db      1239 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298

```

```

QY      381 SerAlaHisLeuAlaValSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIlySg 400
Db      1299 TCGGCCCACTGGGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1358
QY      401 LysHisIlyHisPheArgAlaIlySgIyllegIuGlnGlnArgIyllegIuAlaProLeu 420
Db      1359 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1418
QY      421 AlaSerGlyIlyProValLeuGlyArgGlnHisLeuArgGlyLeuProAlaLeuValAla 440
Db      1419 GCCTCAGAGTGGCTGCTGCTGGGCGGAGACACTGGGAGAGTGGCGCTGTGTG 1478
QY      441 TyrluSerGlnAlaGlyIlyProValGlnArgHisGlnHisPheHisGlnHis 460
Db      1479 TATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1538
QY      461 HisHisIleThrHisPheTyrluThr 470
Db      1539 CACCACTTACCAACCACTTACCAAGACA 1568

RESULT 3
US-09-993-966-5
; Sequence 5, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-993-966-5

Alignment Scores:
Pred. No.: 4,86e-227 Length: 1859
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-993-966-7 (1-470) x US-09-993-966-5 (1-1859)
QY      1  MetGlyLysLeuHisSerIlySgIyllegIuAlaValCylyAargIlySgIyllegIuIuTPIllegIy 20
Db      224 ATGGGAACTTCACTCCAAAGCCGCGCTGTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAG 283
QY      21  AapSerPheAlaValSerAlaAlaATPAlaArgIlySgIyllegIuIuTPIllegIyArg 40
Db      284 GACAGCTTGCGCTGAGAGCGCTGCTGGGCTCGGAAGGCAATCGAGAGTGGATCGGAGA 343
QY      41  GlnArgCySerProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db      344 CACGCGTGGCGCGGCGGTGTCTGGGACCCCGACAGCTGGCGTGGCGGCAACATAGGC 403
QY      61  ArgSerThrArgGlyLeuValGlyAapValLeuArgAphThrLeuSerGlnGluGlu 80
Db      404 CGAAGACCCCGGAGAGCTCGTGGGCGAGAGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAG 463
QY      81  AapAapPheArgLeuGluValAlaLeuProProGlyIleThrAapGlyLeuGlySerGly 100

```

```

Db      464 GACGACCTTGGCTGGAGAGTGGCCCTGCTCTGAGAAAGCTAGACGGGCTGGGACGGGA 523
Qy      101 AspGluLysLysMetGluArgValSerGluProCysProGlySerLysGlnLeuLys 120
Db      524 GATGAGAAAGAAAGTAGAGAGTAGAGAACTCTGCCAGAGCTTCAAGAAAGACGCTGAG 583
Qy      121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db      584 TTGGAAGAGCTCAAGTCCAGCGTCTCATGAGAGAGACCGGCGAGGTGACCTTC 643
Qy      141 ThrLeuTyrAspPheAspAsnGlnLysValThrArgLysAspLethrSerLeuLeu 160
Db      644 ACCCTGATATACCTTGACCAACAGGAGAGCTACCCGAGAGACATCACAGCTTCTG 703
Qy      161 HisThrLethrGluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db      704 CACACCATCTATGAGGTGGTGGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATG 763
Qy      181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysValAspSerValLeuVal 200
Db      764 CTCGGGTAAAGCTCAACCGTGGCCCGATGGCAGCCAGCAAGAGAGACGCTCTGTG 823
Qy      201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnThrLysProThrGluAspLeu 220
Db      824 ATTCAGGCTACCTGCGAGAGCGCAGGCGCCGAGCAGAGACCAAGCCCACTGAGAGACTG 883
Qy      221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db      884 CGAGAGCTGGAGAAAGAGAGAGCGAGCGCCGCTCAGGTTCACAGGGGTGACAGCGCGCTGAG 943
Qy      241 GlnSerGlyCysTrpThrHisCysValAspGluAsnLethrGluArgAsnHisThrLeu 260
Db      944 CAGTCTGGCTGCTACCAACATGGTAGTAGAGAAACATCAAGAGAGAAACCACTACTTA 1003
Qy      261 AspLeuAlaGlyLethrGlnAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db      1004 GATCTCGCCGGGATGAAAACCTACACGTCCTCAATTGGGCTGGCTCCCTTCCTGGGCC 1063
Qy      281 GlnLysSerGluLeuProArgTrpThrSerAsnProThrArgSerArgSerHisGlnPro 300
Db      1064 CAGAGAGCAAGACGCCCCCGCACCTCCATCCCATCGATCTCGCTCCCAAGAACCGG 1123
Qy      301 GlnAlaLethrHisLeuProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db      1124 GAAGCCATCCACATCCACACCGAAAGCCCAAGCGGTGAGCCGCGCTCTTCACCTTC 1183
Qy      321 LeuAspThrProLethrAlaLysValSerGlnLeuGlnArgLeuArgGlyThrGlnAsp 340
Db      1184 CTGGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGGGCAACCAAGAC 1243
Qy      341 GlySerLysHisPheValArgSerProLysValGlnGlyLysSerValGlyValGlyHis 360
Db      1244 GGGAGCAAGACCTTGTAGAGGTCCCCCAAGGCCCAAGGAGTAGTGGGTGGGCGAC 1303
Qy      361 ValAlaArgGlyValArgAsnLysProProLeuGlyProAlaLethrProAlaValSerPro 380
Db      1304 GTGGCCAGAGGGGCAAAACAAAGCCCTCTGTGGACCCCGCATCCCTGCGGTCTCCCC 1363
Qy      381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1364 TCCGCCCAACCTGGTGGCAGCGCCGCTCTCCCTCCCTTACGCCCCCTCGGGCAAG 1423
Qy      401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      1424 AAGGACAAGACACGAGCCAAAGAGAGCAGAGAGGAGTGGCGGGGCTGACAGGACCACTG 1483
Qy      421 AlAserGlyGlyProValLeuGlyArgGlnHisLeuArgGlyLeuProAlaLeuValVal 440
Db      1484 GCTTCAGGTGGCTTGTCTGGGGGAGAGACCTTGGGAGCTGGCCGCTTGTGTGTG 1543
Qy      441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGlnHisGlnHisGlnHis 460

```

```

Db      1544 TATGAGAGCCAGGCGGGGAGCGGCTCAGAGACATGAGCACCAACCATGAACAT 1603
Qy      461 HisHisHisThrHisSerPheTyrGlnThr 470
Db      1604 CACGACCATTTACCAACCATTTTACAGAC 1633

RESULT 4
US-09-993-966-2
; Sequence 2, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NRK PROTEIN,
; TITLE OR INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USBS THEREOF
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-2

Alignment Scores:
Pred. No.: 7,79e-198 Length: 1416
Score: 2189.50 Matches: 409
Percent Similarity: 91.06% Conservative: 19
Best Local Similarity: 87.02% Mismatches: 41
Query Match: 87.62% Indels: 1
DB: 10 Gaps: 1

US-09-993-966-7 (1-470) x US-09-993-966-2 (1-1416)
Qy      1 MetGlyLysLeuHisSerLysProAlaAlaValCysLysArgArgGluSerProGlnGly 20
Db      1 ATGGGGGAACTTCACTGAGAGCCGGCCGCTGGCAAGCGCAGAGGAGAGCCCGAAGGT 60
Qy      21 AspSerPheAlaValSerAlaAlaTrpAlaArgLysGlyLethrGluTrpLethrArg 40
Db      61 GACAGCTTCTGTAAGCGCTGCTGGCAAGAAAGCATCGAGAGTGGATCGGAGG 120
Qy      41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrLethr 60
Db      121 CAGGCTGTCAAGGACCGTCTCAGACCCCTGACGCTGAGATTGGCAGGACCTGTTGCT 180
Qy      61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGlnGluGlu 80
Db      181 CGAGGCACTGGGGAACGCGGGGTGACATTTAGAAAGGCTCTCGGTGAGAGACGAG 240
Qy      81 AspAspPheArgLeuGlnValAlaLeuProProGlnLysThrAspGlyLeuGlySerGly 100
Db      241 GACGACCTTCCCTTAGAAGTGGCCCTGCGCTCGAGAAAGATCAACACCTAAGTAGTGA 300
Qy      101 AspGluLysLysMetGluArgValSerGluProCysArgProGlySerLysGlnLeuLys 120
Db      301 GATGAGAAAGAAAGTAGAGAGCTGAGGAACTGGCCAGGCTTCCAAAGAAAGACCTAAG 360
Qy      121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db      361 TTGGAAGAGCTACAGTGTGATGTCTGTGAGAGAGACAGCGGCAAGAGTGAACCTTTC 420
Qy      141 ThrLeuTyrAspPheAspAsnGlnLysValThrArgLysAspLethrSerLeuLeu 160
Db      421 ACTCTATATACCTTGCACCAACATGGCAAGTAGACCCGTGAGGACATTACCAAGCTTCTG 480

```

```

QY 161 HistHrLleTyrGluValAlaPheSerSerValAlaMetHisSerProThrSerSerLeuMet 180
DB 481 CATACCTCTATGAAAGTGTGACTCTCTGTGAACCTTCCCACTCAAGCAAGCA 540
QY 181 LeuArgValIleuThrValAlaProAspGlySerGlnSerIleValArgSerValLeuVal 200
DB 541 CTGGGGGTGAAGCTCACGGTGGCTCTGACGGGAGCCAGATTAGAGAGGTCTTTTC 600
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleValProThrGluAspLeu 220
DB 601 AACCATACCATCTGCAGAGACAAAGGCCCGGACAGACACCAACCCGCTGAGAGACTG 660
QY 221 ArgSerTrpGluValIleValGlnArgAlaProLeuArgPheGlnIleValAspSerArgLeuGlu 240
DB 661 CGTGGCTGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 241 GlnSerGlyCysTrpHisIleCysValAlaAspIleValIleGluValGluValIleVal 260
DB 721 CAGCGACACTGCTACCACTTGGCTGATGAGAACATTGAGAGAGAGAGAGAGAGAGAGAG 780
QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
DB 781 GACCTGGCGGGATGAGAACTACACGCTCAGATTGAGACCGGATCCCTTCGGTGGCC 840
QY 281 GlnIleSerGluLeuProProArgTrpSerAspProThrArgSerArgSerHisGlnPro 300
DB 841 CAGAGAGTCAAGAGCTGCCCCCTCGAATCTCAACCCCTGCTCTCCCTCCCAAGAGCA 900
QY 301 GlnAlaIleHisIleProHisArgIleProGlnIleValAspProAlaSerPheHisPhe 320
DB 901 GAAGCTCCCAATCCCAACCGAGGCGCCCAAGGTGTGACCCAGGCTCTTCCACTC 960
QY 321 LeuAspThrProIleAlaIleValSerGluLeuGlnIleValArgLeuArgIleValThrIle 340
DB 961 CTGGACACCCCATTTGGCCAAAGCATCAGAGCTCAGACAGCGCTCCGGGCACTCAGAGAT 1020
QY 341 GlnSerTrpHisPheValArgSerProValAlaGlnIleValSerValGlyValGlyHis 360
DB 1021 GGGAGCAAGACATTGTGAGGTCCCGCAAGGCCCAAGGCAAGACATGGGTATGGCCAC 1080
QY 361 ValAlaArgGlyAlaArgAsnIleValProLeuGlyProAlaIleProAlaValSerPro 380
DB 1081 GGGGCGCAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuIleProSerIleValAlaProLeuGlyHis 400
DB 1141 TCTGCCCATCTGGCCACCAAGCCCTTCTCCCACTGGGACCCCTGGGAGCAAG 1200
QY 401 IlyHisIleValHisArgAlaIleValSerGlnIleValArgIleValLeuGlnAlaProLeu 420
DB 1201 AAACACAGCATGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 421 AlaSerGlyIleValProValIleGluValArgIleValIleValArgIleValAlaLeuVal 439
DB 1261 GCTGCAGAGAGCTCCACCTCATGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 440 ValTrpIleSerGlnAlaGlyIleProValGlnArgIleGluHisIleHisIleGlu 459
DB 1321 GTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 460 HisIleHisIleHisIleHisIleHisIleHisIleHisIleHisIleHisIleHis 469
DB 1381 CATCACCACTATATACCACTTCTATCAG 1410

```

```

; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 6
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Mus sp.
; US-09-993-966-6

Alignment Scores:
Pred. No.: 3,91e-196 Length: 1401
Score: 2171.50 Matches: 406
Percent Similarity: 91.01% Conservative: 19
Best Local Similarity: 86.94% Mismatches: 41
Query Match: 86.89% Indels: 1
DB: 10 Gaps: 1

US-09-993-966-7 (1-470) x US-09-993-966-6 (1-1401)
QY 1 MetGlyIleValLeuHisSerIleValProAlaIleValCysIleValArgIleValSerProGlnIle 20
DB 1 ATGGGGAAATCTTCACTGAAAGCCGGCCGCTGTGCAAGCGAGAGAGAGAGAGAGAGAGAG 60
QY 21 AspSerPheAlaValSerAlaAlaIleValAlaArgIleValIleGluIleValIleValArg 40
DB 61 GACAGCTTCTCTGAAAGCGCTGTGGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 41 GlnArgCysProGlyIleValIleSerGlyProArgGlnLeuArgIleValIleValIleGly 60
DB 121 CAGCGCTGTCCAGGACAGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 61 ArgSerThrArgIleValIleValIleValIleValIleValIleValIleValIleValIle 80
DB 181 CGAGGCACTGGGAACTCGTGGGTGACATTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 81 AspAspPheArgIleGluValAlaIleValProGlnIleValThrArgIleValIleValSerGly 100
DB 241 GACGACTTCCCTTGAAGAGGCGCTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 101 AspGlnIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 120
DB 301 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 121 PheGlnIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 140
DB 361 TTGGAAGAGTCAAGTGTGATGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 141 ThrLeuTrpAspPheAspAlaIleValIleValIleValIleValIleValIleValIle 160
DB 421 ACTCTATATGACTTGCAGCAACATGGCAAGTACCCGTGAGAGAGAGAGAGAGAGAGAG 480
QY 161 HistHrLleTyrGluValAlaPheSerSerValAlaMetHisSerProThrSerSerLeuMet 180
DB 481 CATACCTCTATGAAAGTGTGACTCTCTGTGAACCTTCCCACTCAAGCAAGCA 540
QY 181 LeuArgValIleuThrValAlaProAspGlySerGlnSerIleValArgSerValLeuVal 200
DB 541 CTGGGGGTGAAGCTCACGGTGGCTCTGACGGGAGCCAGATTAGAGAGGTCTTTTC 600
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleValProThrGluAspLeu 220
DB 601 AACCATACCATCTGCAGAGACAAAGGCCCGGACAGACACCAACCCGCTGAGAGACTG 660
QY 221 ArgSerTrpGluValIleValGlnArgAlaProLeuArgPheGlnIleValAspSerArgLeuGlu 240
DB 661 CGTGGCTGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720

```

QY 241 GlnSerGlyCysTyrHisIscysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
 DB |||
 Db 721 CACCCAGACGCTTACCAACCAATGGTGATGAGAACTTGAAGAGAAACCACTACTTA 780
 QY 261 AspleuAaGlyIleGluAenTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 Db 781 GACCTGGCGGGATGAGAACTACACCTCTCAGTTTGAACCGGATCTCCCTCGGTGGCC 840
 QY 281 GlnIysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
 Db 841 CAGAGTCAGAGCTGCCCCCTCGAATCTCCAAACCCCACTGCTCTGCTCCACAGACCA 900
 QY 301 GluAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320
 Db 901 GAAGCTCCCAACATCCCAACCGGAGGCCCAAGGTGTGACCCAGGCTCTTCCACCTC 960
 QY 321 LeuAspThrProIleAlaIysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
 Db 961 CTGGACACCCCATTTGGCCMAAGGCATCAGAGCTCCAGACGGCTCCGGGCACTCAGAT 1020
 QY 341 GlySerIysHisPheValArgSerProIysAlaGlnGlyIysSerValGlyValGlyHis 360
 Db 1021 GGGAGCAAGACATTGTGAGGTCCTCCCAAGGCCCAAGGCAAGACATGGGTATGGGCAC 1080
 QY 361 ValAlaArgGlyValArgAsnIysProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 1081 GGGGGCCAGAGGTGCMAAGAGCACCTCCACTGGTACCACACCCCACTACTGTTCCCC 1140
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
 Db 1141 TCTGCCATCTGGCCACACGCCCCCTTCTCCCACTGGGACCCCTGGGGCAACG 1200
 QY 401 LysHisIysHisArgAlaIysGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 Db 1201 AAACACAGACATCGAGCCAAAGAGACAGGCGAGCTGCCGGGCTGCAAGGCCCTCG 1260
 QY 421 AlaSerIleGly---ProValLeuGlyArgGluHisLeuArgGlyLeuProAlaLeuVal 439
 Db 1261 GCTGCAAGAGGCTCCACCGTATGGGCGGAGAGAGAGGAGGCTGCTGGCTGGGTG 1320
 QY 440 ValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisGlu 459
 Db 1321 GTGTACAGAGCCAGGCTGGGCAAGCGCTCCAGAGACACGACATCACCAACCA 1380
 QY 460 HisHisHisHisTyrHisHis 466
 Db 1381 CATCACCACTATCACAC 1401

RESULT 6
 US-09-730-989-1
 ; Sequence 1, Application US/09730989
 ; Patent No. US20020061552A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yan, Dong
 ; TITLE OF INVENTION: MAMMALIAN DISHEVELLED-ASSOCIATED PROTEINS
 ; FILE REFERENCE: PP-01657.002 / 200130.518
 ; CURRENT APPLICATION NUMBER: US/09/730.989
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PasteSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1401
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-09-730-989-1

Alignment Scores:
 Pred. No.: 9,38e-196 Length: 1401
 Score: 2167.50 Matches: 406
 Percent Similarity: 90.79% Conservative: 18
 Best local Similarity: 86.94% Mismatches: 42

Query Match:	86.73%	Indels:	1
DB:	9	Gaps:	1
US-09-993-966-7 (1-470) x US-09-730-989-1 (1-1401)			
QY 1 MetGlyLysLeuHisSerIysProAlaAlaValCysIysArgArgGluSerProGluGly 20			
Db 1 ATGGGAAATTTACTCTGAAGCCGGCGCGCTGTGCAAGCGGAGGAGGAGCCCGAAGGT 60			
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgGlyGlyIleGluGluTTPIleGlyArg 40			
Db 61 GACAGCTTGGCTTAAGCGCTGCTGGCAAGAGAAAGGCATTCAGAGGTGAATCGGAGG 120			
QY 41 GlnArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60			
Db 121 CAGGCTGTCAGACGACCGTCTTCAAGACCCGCTCAAGCTGATGATTGGCAGGCACTGGT 180			
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80			
Db 181 CAGAGCACTCGGGAACCTCGTGGGTGACACTTCAAGAGGCTCTCGGTGAGAGAGCAG 240			
QY 81 AspAspPheArgLeuGluValAlaLeuProProGluIysThrAspGlyLeuGlySerGly 100			
Db 241 GACGACTTCCCCCTAGAGAGTGGCCCTGCCCTGAGAGAGATCGACAGCTTAGGTATGGA 300			
QY 101 AspGluLysLeuMetGluArgValSerGluProCysSerProGlySerIysGlnLeuLys 120			
Db 301 GATGAGAGAGAGATGAGAGACTGAGGAACTGGCCAGACTGGCCAGGCTTCCAGAGCACTCAG 360			
QY 121 PheGluGluLeuGlnCysAspValSerMetGluLysAspSerArgGlnGluTTPThrPhe 140			
Db 361 TTGGAAGACTTACAGTGTGATGTCTGTGGAAGAGACAGCCGGCAAGGTGACTTTC 420			
QY 141 ThrLeuTyrAspPheAspAsnGlnGlyValThrArgGluAspIleThrSerLeuLeu 160			
Db 421 ACTCTATATCACTTCGACCAATGGCMAAGTACCCGTAGAGACATATACACTTGGCTG 480			
QY 161 HisThrIleTyrGluValValaAspSerSerValAsnHisSerProThrSerSerIysMet 180			
Db 481 CATACCACTATGAGAGTGTGACTCTCTGTGAACATTTCCCACTACAGACAGACA 540			
QY 181 LeuArgValIysLeuThrValAlaProAspGlySerGlnSerIysArgSerValLeuVal 200			
Db 541 CTGGGGTGAAGCTCACCGTGGCTCTGACGGGACAGAGTAAAGAGAGCGTCTTTC 600			
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeu 220			
Db 601 AACCATACCGATCTGCAGACACAGAGCCCGGAGCAGACCAAAACCGCTGAGAGACTG 660			
QY 221 ArgSerThrGluLysIysGlnArgAlaProLeuArgPheGlnIysAspSerArgLeuGlu 240			
Db 661 CTGGCTGGAG 720			
QY 241 GlnSerGlyCysTyrHisIscysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260			
Db 721 CACCCAGACGCTTACCAACCAATGGTGATGAGAACTTGAAGAGAAACCACTACTTA 780			
QY 261 AspleuAaGlyIleGluAenTyrThrSerGlnPheGlyProGlySerProSerValAla 280			
Db 781 GACCTGGCGGGATGAGAACTACACCTCTCAGTTTGAACCGGATCTCCCTCGGTGGCC 840			
QY 281 GlnIysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300			
Db 841 CAGAGTCAGAGCTGCCCCCTCGAATCTCCAAACCCCACTGCTCTGCTCCACAGACCA 900			
QY 301 GluAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320			
Db 901 GAAGCTCCCAACATCCCAACCGGAGGCCCAAGGTGTGACCCAGGCTCTTCCACCTC 960			
QY 321 LeuAspThrProIleAlaIysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340			
Db 961 CTGGACACCCCATTTGGCCMAAGGCATCAGAGCTCCAGACGGCTCCGGGCACTCAGAT 1020			

Oy		440	VaLTYrclUsErGlnAagLYgInPcovaLGlnARHIGlWHIsHIsHISglU	459
Dd		1263	GtTACAGAGGCCAGCTGGGCGAGCCGCTTCAGAGACAGAACCATCAACACGAA	1322
Oy		460	HIsHIsHIsTyRHIsHIsPhetyRgin	469
Dd		1323	CATCACCATTTATCACCACCTCTATCAG	1352
		RESULT 8		
		US-10-087-192-178		
		/ Sequence 178, Application US/10087192		
		/ Publication No. US20020182586A1		
		/ GENERAL INFORMATION:		
		/ APPLICANT: Morris; David W.		
		/ APPLICANT: Engelhard, Eric K.		
		/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR		
		/ FILE REFERENCE: 52945200122		
		/ CURRENT APPLICATION NUMBER: US/10/087,192		
		/ CURRENT FILING DATE: 2002-03-01		
		/ PRIOR APPLICATION NUMBER: US 09/747,377		
		/ PRIOR FILING DATE: 2000-12-22		
		/ PRIOR APPLICATION NUMBER: US 09/798,586		
		/ PRIOR FILING DATE: 2001-03-02		
		/ NUMBER OF SEQ ID NOS: 2059		
		/ SOFTWARE: FastSeq for Windows Version 4.0		
		/ SEQ ID NO 178		
		/ LENGTH: 106315		
		/ TYPE: DNA		
		/ ORGANISM: Homo sapiens		
		/ FEATURE:		
		/ NAME/KEY: misc_feature		
		/ LOCATION: (1)..(106315)		
		/ OTHER INFORMATION: n = A,T,C or G		
		US-10-087-192-178		
		Alignment Scores:		
		Pred. No.:	9.25e-101	Length: 106315
		Score:	1191.50	Matches: 315
		Percent Similarity:	26.36%	Conservative: 1
		Best Local Similarity:	26.27%	Mismatches: 1
		Query Match:	47.68%	Indels: 883
		DB:	13	Gaps: 3
		US-09-993-966-7 (1-470) x US-10-087-192-178 (1-106315)		
Oy		154	GlUaSpIleThrSerLeuLeuHIsThrlEtyRGluValAlaSpSerSeValAenHis	173
Dd		92720	CAGAHCATCACCGACTTGCTGCACACCATTAAGGTGTGAATCTCTCTGTCAACAC	92779
Oy		174	SerProThrSerSerLySMetLeuArgVallyLeuThrValAlaProAspGlySerGln	193
Dd		92780	TCCCCAACATCCAGACAAGATGCTCGGGGTAAAGCTCACCGTGGCCCCGATGGACGCCAG	92839
Oy		194	SerLySaTrSerValLeuValAenGlnAla-	203
Dd		92840	AGCAAGAGGAGCGCTCTGTTCATCAACAGCGTGTAGAGGCGTCAGGGCTGAGCCTGGAAA	92899
Oy		203	-	203
Dd		92900	CAATGTCTCTCCATCTCAGAGAGACAGACCTTCTCTGGAGGGCCGGAGAGGAACCTG	92959
Oy		203	-	203
Dd		92960	TGGGCTTTCTGGGGCAGCTGCTTGTGTGACAGAGTGGGTTCAATGGGTCTTGAAGGG	93019
Oy		203	-	203
Dd		93020	TAGAGGAAAGGTGGGAAGGAGGAGGAAGATTCTGGTGAAGAGCTCATATAAGCAG	93079
Oy		203	-	203
Dd		93080	AGCCATCGACGTGAGTAGAAGCGTGTGTTACGGGACGATGGCGCTCTGCACACCTTGAG	93139

QY	203	-----	203	-----	203
Db	93140	GCCAGCTGAAGAGGTGAGCTTGAGGGGGAGTCTGGGGGAGAGGTGAGGCTAGATGG	93199	-----	93258
QY	203	-----	203	-----	203
Db	93200	GGATGTGTCCAGGGCAGGTAGCCTTGAGACCCCTTATTTAGCCCACTCAGGAAACC	93255	-----	93314
QY	203	-----	203	-----	203
Db	93260	TCCCACTGCGCTTGCAAAAGACGTGCACGCCAGGACGCTGGGGAGGTGGAGCAG	93319	-----	93378
QY	204	-----	AspLeuGlnSerAlaArg	209	-----
Db	93320	GCAGCTCACCTGGTGTCTCCTGCTTCTCGGGCCGGACTCAGACCTTCAGAGCCGCAAG	93379	-----	93443
QY	210	ProAlaGlnIleThrIysProThrArgIleAspLeuArgSerTyrGlnIleIleArgAla	229	-----	93498
Db	93380	CCCCGAGCAGACCAAGCCCACTGAGACCTGCGGAGCTGGAGAAAGACAGCGAGCC	93433	-----	93500
QY	230	ProLeu-----	231	-----	231
Db	93440	CCGCT--CAGGTATGTGGGCAATGTGCATGAGCATATGTTGAGCACCAAGCTGTGACCT	93499	-----	93558
QY	231	-----	231	-----	231
Db	93499	GCTTCTGAAGATTGAGAGCAGAGAGTTTGTGCTGTGTGTCTCTTCAGAGGACCA	93558	-----	93618
QY	231	-----	231	-----	231
Db	93559	ACACTTTTGATGCACAGACTCATAAATCAATGAAATCTTAACAATAATACATGTTAG	93618	-----	93678
QY	231	-----	231	-----	231
Db	93619	TCCACATCTTTTATGTGAACACATCTTGGCAGACAGACTGGAGGTGACATTTTGTGG	93678	-----	93738
QY	231	-----	231	-----	231
Db	93739	AGATGACTGAATTTGCACACTAGTGAGATTAAGAAGCGTAAACACTTCCATCCCTTAA	93739	-----	93799
QY	231	-----	231	-----	231
Db	93799	GGGTGGGCTGTGCTGCCACGTGGTTATGAAACTCTAGTTTTCAGGCTTTTGTGACTTT	93855	-----	93915
QY	231	-----	231	-----	231
Db	93859	CAGAAATGTGGATTAAGGAATTATGAGGCTGTAGTTACCTTTACCTGTACTTCTTCAGC	93915	-----	94000
QY	231	-----	231	-----	231
Db	93919	TGCAAGGAGGAAGGAGGAAAGAAAGAGAGATTGTTAAAGAAAGGATTGATTAAC	93978	-----	94038
QY	231	-----	231	-----	231
Db	93979	AACAGTAACAGGCGAGGATTAAGGCTGTGTTTAACTCTAGTGTCAAAATGTTAAGGAAT	94038	-----	94099
QY	231	-----	231	-----	231
Db	94039	GATCGAGCTGGAATTAGTACTGCTTAGGGGAAAGAAAACAGTTCAACGAGGCTCACAG	94099	-----	94159
QY	231	-----	231	-----	231
Db	94099	CCAGTGTGGGACAAGGGCTGAAGATCAGAGGACCAGGCTGGCTTAGAGATCTGGGGG	94159	-----	94219
QY	231	-----	231	-----	231
Db	94159	AAGGTGTACTGTGTGTGTGGGAAGGAACCGAGAGGAGCAGATCGAAGAGGTCTTGG	94219	-----	94279

QY 231 ----- 231
 Db 94219 GAGAGAGGGCTTAGGAGCCCCCGAGCCCTGCAGGGGAAAAAAGAGCCTTGAGCATC 94278
 QY 231 ----- 231
 Db 94279 TCTCCCAACCCCTGCATGACACTGTGTAAAGTCAACATGATGTGACACACACTCA 94338
 QY 231 ----- 231
 Db 94339 CACACTGTGAGGCCCATCTGGGCGTAGGGCAGTCAATGGGTCTGCTGGCCAAAC 94398
 QY 231 ----- 231
 Db 94399 CCAACAGTTTAAAGGGATTCTGCTGCCACACCCCTGAGAGACAGGCCAGGGTTAGG 94458
 QY 231 ----- 231
 Db 94459 CCGGCTACCTGCAGTGGGCATATATGGCTTGAAGCCTGGCACTGGGGCATTTCTAGAACT 94518
 QY 231 ----- 231
 Db 94519 TGACCAATACTCTCACTTTGCTTTGAACCTGTTCAGACAAAACCCGTCACACTCT 94578
 QY 231 ----- 231
 Db 94579 CCCAGAAAGCTGCGGGAGGTGCGGCTGTGGCTGTGTCTATGTCTGTCTTCATCGG 94638
 QY 231 ----- 231
 Db 94639 CAAGAGGGAATGAGGACACAGTTCGTATAGAGACTGTTCCCTCCCACTCTCCAAA 94698
 QY 231 ----- 231
 Db 94699 GGCCAAAGAAATGATGAGAGGCTCCAGAGTTCAATTCTGGGGCTTCTAGAGCTA 94758
 QY 232 ----- Arg 232
 Db 94759 TGGCTTGGCCCCCACTGTGTGTGTATCCCACTAGCTGCCCTGCCGTGAGG 94818
 QY 233 PheGlnGlyAspSerArgLeuGlnSerGlyCysTyrHisCysValAspGluAsn 252
 Db 94819 TTCAGAGGTGACAGCCCGCTGAGACAGTCTGGCTGCTACCACTTCGTAGATGAGAAC 94878
 QY 253 IleGluArgHisIleuAspLeuIleGlyTyrHisSerGlnPhe 272
 Db 94879 ATCGAGGAGGAAACCACTATTAGATCTCGCGGATAGAAAACTAACAGTCCCAATTT 94938
 QY 273 GlyPro----- 274
 Db 94939 GGGCTGTGTAAGGACTCAAGCAACCTGCAATGGCGATGAGGCAAGGCGTGGAC 94998
 QY 274 ----- 274
 Db 94999 GGGCCAGGGGGCGGTGGGGGTGTTCACCTGACCCAGGCTTGCTGAAGACACTA 95058
 QY 274 ----- 274
 Db 95059 TTATGGAGACAGTCAAAAGACTTGTGAGAAAGTAGCTTAAAAATGTTCAAAATTA 95118
 QY 274 ----- 274
 Db 95119 AGTAATACATGACGAATTCGAACAAATTCMAATAGTCTAAAGGCTTAGCAACCAATG 95178
 QY 274 ----- 274
 Db 95179 ACAAAACAAGTTCTGTGCCCCATCCAGGCCAACCAAGCCGTGTTCCGGGGCAACAC 95238
 QY 274 ----- 274
 Db 95239 TTGACCTCTCTGCTGTTCTTCCAGAAAGACTCGGTTTCCCTAAGTACAGACTA 95298
 QY 274 ----- 274

Db 95299 CTGTCTTTATACGCTTGCATATGTGACCTCTTGCTCAATTTCACTCCCAACTC 95358
 QY 274 ----- 274
 Db 95359 CTTTCCCTTTCTCTCATATGTTTAATTCCTATAGTACTTCTGATGTTTTA 95418
 QY 274 ----- 274
 Db 95419 GCATATACTTAGAGTCTTTCTTGCCCAAGTACCTTAGATAGTTTCGGGGTCTGAC 95478
 QY 274 ----- 274
 Db 95479 TGCCATGTCTTCTTCGAGCTCCCTCGAGAGATTTTGAAGACTTCAAGAAC 95538
 QY 274 ----- 274
 Db 95539 AGTTCTTCCCATAGAGATGGGAAAGTTCCATTTGAGAGATTTCTGCAAGGAGT 95598
 QY 274 ----- 274
 Db 95599 GAGGGCAGTCAAGGCAATGGGGGTAGCTCTGGTTTGGAGAACTGGGGCGGGGTGA 95658
 QY 274 ----- 274
 Db 95659 TGTCTGGGTATAGCGCAACCCAGACACAGTAGGTCTATTAATGTCTGTGAAT 95718
 QY 275 -----GlySerProSerValAlaGlnLeuSerGlyIleuProProArgHisSerAsn 291
 Db 95719 TGGTCTTAGGCTTCCCTTCGTGGCCAGAACTGACACTGCCCCCAGCTCCAT 95778
 QY 292 ProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLeuProGln 311
 Db 95779 CCCACTGATCTGCTCCCATAGAGCCGGAAGCATCCACATCCACACGAAAGCCCCAA 95838
 QY 312 GlyValAspProIleSerPheHisPheLeuAspThrProIleAlaValSerGlyIleu 331
 Db 95839 GGGGTGACCCCGCTCTCTTCACTTCTTGACACCCCAATCCGCAAGTCTCAGACTC 95898
 QY 332 GlnGlnArgLeuArgGlyThrGlnAspGlySerLeuHisPheValArgSerProValAla 351
 Db 95899 CAGCAACGCTCCGGGGACCCAGACGAGAGGAGACACTTGTGAGTCCCCCAAGGCC 95958
 QY 352 GlnGlyValSerValGlyValGlyHisValAlaArgGlyAlaArgLeuProProLeu 371
 Db 95959 CAGGGCAAGAGTGTGGGTGGGCAAGTGGCCAGAGGGCAAGAAACAAGCCCTCTG 96018
 QY 372 GlyProAlaIleProAlaValSerProSerLeuHisLeuAlaIleSerProAlaLeu 391
 Db 96019 GAACCGCATTCCTCGGTGTCCTCCCTCGCCACCTGTGCTCCAGCCCGGCTTCCTC 96078
 QY 392 ProSerLeuAlaProLeuGlyHisIleValSerHisIleValArgAlaValArgLeuGln 411
 Db 96079 CCTTCTAGCCCCCTCCGGGACCAAGACACAGCAACGACCAAGAGAGACCAAG 96138
 QY 412 GlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyValProValLeuGlyArgGlnHis 431
 Db 96139 GGTGCGGGGCTGAGGACCACTGAGTGGCTGAGTGGCTCTCGGGGGGAGAC 96198
 QY 432 LeuArgGlyLeuProAlaLeuValIleTyrGlnSerGlyAlaGlyIleProValGlnArg 451
 Db 96199 CTGGGAGCTGCGCCCTTGTGTGTATGAGGACGAGCGGGGAGCGGGTCCAGAGA 96258
 QY 452 HisGluHisIleHisIleGluHisIleHisIleTyrHisIlePheTyrGlnThr 470
 Db 96259 CATGAGCACACACACATGAATCAACCACTTACCAACCACTTACCAAGACA 96315

RESULT 9
 US-10-087-192-175
 ; Sequence 175, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.

APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE OF INVENTION: CANCER
FILE REFERENCE: 52945200122
CURRENT APPLICATION NUMBER: us/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIORITY APPLICATION NUMBER: US 09/747,377
PRIORITY FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 175
LENGTH: 98474
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(98474)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-175

Alignment Scores:
Pred. No.: 7.57e-90 Length: 98474
Score: 1076.00 Matches: 278
Percent Similarity: 29.11% Conservative: 18
Best Local Similarity: 27.34% Mismatches: 39
Query Match: 43.06% Indels: 683
DB: 13 Gaps: 6
US-09-993-966-7 (1-470) x US-10-087-192-175 (1-98474)

QY 129 SerMetcGluaPserAArgIngluTrpThreHleuTyraPheAspAsn 148
Db 83990 TCTCAGGTGCTGATGCGAGGATCGTG-----CTAACTCTTG 84031
QY 149 G1YlyVal-----ThArgGluAsp1LeHserLeuH1sThr1LeTyG1uVal 166
Db 84032 GGGGTCTGCTCTCCCAACAGACATTCACAGCTTGCTGCACTACATCTATGAAAGTG 84091
QY 167 ValAspSerValaAmH1sSerProThSerSerLyMetLeuArgVal1yLeuThr 186
Db 84092 GTTACTCTCTGTGAACATTCCTCCCAATCAAGACACACTGCGGTGTAAGCTCACCC 84151
QY 187 ValAlaProAspG1ySerGInserLyArgSerVal1euValaAm----- 201
Db 84152 GTGGCTCCTGACGGAGCCAGATGAAGAGAGCGCTCTTTCACATACCGGTAAAGG 84211
QY 201 ----- 201
Db 84212 GCGCGAGACGTGCCACATATGGAAAGACAGGAGATCTAAGTCAGCTGGAGGGTG 84271
QY 201 ----- 201
Db 84272 GGAAGACCCAGTACATAGACAGTTCCTGTGTGTACTGCAAGGCTTCAGCAA 84331
QY 201 ----- 201
Db 84332 ACCCTTGAAGACTAGGCTCTGGGAGGAGAGGAGCGGAATCACAGATGAAAGG 84391
QY 201 ----- 201
Db 84392 CCCAGTGAACAAAGCATAGATTGAGACTGAGAGAGATGACGCGCAGCTTAAGAG 84451
QY 201 ----- 201
Db 84452 GCAGCTTATGAACTGAGAGGTGTGGGACAGGCAACTAGTTTGCTCTTGCTGGCA 84511
QY 201 ----- 201
Db 84512 ACCCTGTGAGAGCTGTGTGAATCTATTGAGTGGCTTTTGAGCACACAATGCCAGGA 84571
QY 202 -----GlnAlaAsp 204

Db 84572 AGCCCTGGCCGAGTTGGAGGAGCACTGACAGTCTCTTGACTTCTGGACTCAGAT 84631
QY 205 LeuGInserAlaArgProAla1aGluThrLyArgProThrGluAspLeuArgSerTrpGlu 224
Db 84632 CTGCAGAGCACAAAGGCCCCGAGGACAGACCAAAACCCGCTAGGAGCTGCTGGGAG 84691
QY 225 LyLeuGInArgAlaProLeu----- 231
Db 84692 AAGAGCAGCGAGCCCACT-CAGGTAATGACATGCTGTGCCCCGTGAACATGATTGAG 84750
QY 231 ----- 231
Db 84751 CGCAGCTGATATGTACATCTGAAAGACTTTGAGTAGAGTTCCATCTGGGTCTCT 84810
QY 231 ----- 231
Db 84811 GCAGGGCCATGATTTATTTTCAGAAAGACAGATTTACAAATCAGACTTAATTACA 84870
QY 231 ----- 231
Db 84871 AACAAAACAGAAAAAAATCAAAACCTAGATATAGCTATCATCTATTCATATA 84930
QY 231 ----- 231
Db 84931 AAATTTCTTAGATTTCTCCATGATGTTGGATTTCAGATGTAATGGTAATTCGTATG 84990
QY 231 ----- 231
Db 84991 TATTTCCATGTAGATAGTAGGTAGGACACTTGGGGGTCTAAACACTGGGATTTCTGC 85050
QY 231 ----- 231
Db 85051 TGGGAGTATGTGATATAGCATTAAGCATCACAGAGGACAAAATTCCAAAGCTCTGTCCC 85110
QY 231 ----- 231
Db 85111 TTAAGATGGCTGGCTGTTCAGTGGCTATGGGAAACCTCCATTTCAGAGTGTCTCAG 85170
QY 231 ----- 231
Db 85171 AATTGTAGTAAAGACTTATAGATCTGCACTTATCTGTCCGTTGTCACTTTCGACATACA 85230
QY 231 ----- 231
Db 85231 CAGAGGAGAGAGGATTTTAAAGAAAGACCTTAACAGACAGAGGAGGAGCAAGCA 85290
QY 231 ----- 231
Db 85291 AGCGTAGATCTTCAACATCTGTGCCAAAGAGCGTGGCAGAGATCGGCTAGCTCTT 85350
QY 231 ----- 231
Db 85351 CATGTAGCGTGTATGAAGACAGACATGTAAGTGGCAGTGGGAGCTGGCTGGAGAT 85410
QY 231 ----- 231
Db 85411 GGGTGGGCTGGGGGCAATTGAACAGACTTGCAGTGTGAAGGAGATGTGTGGGCACT 85470
QY 231 ----- 231
Db 85471 GGGGAATGAAGGACAGACAGATTCGTGCTCAGACGCTTCTCCAAATGCACAGAA 85530
QY 231 ----- 231
Db 85531 ACGTGAAGATGATGAAGCTCCGCGTTTGAGCTTATCCAGACAGGCTCAGAGGT 85590
QY 232 -----ArgPheGlnGlyAspSerArgLeuGlu 240
Db 85591 CTGTGCTTCTTACATATATGCTCTTTTGTGCACAGGTTCCAGGGTGACAGCACTGGAG 85650
QY 241 GInserG1ySerTyRHisH1sCyvValAspGluAsn1LeGluArgArgAsnH1sTyRLeu 260
|||

Db 85651 CAGCCAGACTGTCTACACACATTCGGCTGGATGAGAACTTGAGAGAGAAACCTACTACTA 85710
Qy 261 Aspleu1a1a1y1leg1u1a1y1ThSerg1nph1e1y1----- 273
Db 85711 GACCTGGCGGGATGAGAACTACACGCTCTCAGTTTGACCGGTAAGGGCTGCTGTGT 85770
Qy 273 ----- 273
Db 85771 GGGCCATGCTTGACACACAGTGTGGTAAGTACCACTGATCCTGAGGACTGGGAG 85830
Qy 273 ----- 273
Db 85831 ATAGTAGCTTTCTACTGACAGGTCAAGGCTGTCTCAGAGCCCTAACAGACACTG 85890
Qy 273 ----- 273
Db 85891 AAGCATTTTCAAAATCAAAATTAATTAATTAAGTGGGCAAAATCCAGAGCAA 85950
Qy 273 ----- 273
Db 85951 GAGAGACTGAGAGTCCAGCAAAAGAGTACCCGCTCAAGCCAGCCCGGAGCC 86010
Qy 273 ----- 273
Db 86011 CACATGACCCCGAGCCCGAGCCCGAGCCCGAGCCCGAGCCCGAGCCCGAGCCCG 86070
Qy 273 ----- 273
Db 86071 AGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCC 86130
Qy 273 ----- 273
Db 86131 GCTGCTTGCAGCCTCTGTGCCCCCTGACCTGTGTGCACTCTCTTCTTAATGTT 86190
Qy 273 ----- 273
Db 86191 GAGCTCTCTGAGTTTCACTGATTCCTTATATTAATTAAGGCTGTGCCACT 86250
Qy 273 ----- 273
Db 86251 CCTTAGTCCCT 86310
Qy 273 ----- 273
Db 86311 CAGAGCTGTAGGAGGATTAGAGGCTGTAGTCTCAGACAGAGTGTCTCTCTGTA 86370
Qy 274 -----Pro 274
Db 86371 ACCCAGGCTAGCCAGTGCACACAGTGGTCTCATGAGTATCTGTGGATTGTCTCT 86430
Qy 275 -GlySerProSerVal1a1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1 294
Db 86431 AGGATCCCTTCGGTGGCCAGAGTCAAGAGTGTGCCCCCTGATCTCCAACTCCAG 86490
Qy 294 gSer1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 314
Db 86491 CTCTGCTCTCCAGAGCAGAGTGTGCCCCCTGATCTCCAACTCCAGAGTGTGGA 86550
Qy 314 pPro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 334
Db 86551 CCGAGGCTCTTCCACTCTTCCAGTGTGAGGCTTCCAGGCACTCAAGGCTCCAGCAAG 86610
Qy 334 gLeu1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1 354
Db 86611 GCTCCGGGACCTCAGAGTGTGAGGCACTTGTGAGGCTTCCAGGCACTCAAGGCTCC 86670
Qy 354 sSerVal1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 374
Db 86671 GAACATGTGGTATGAGGCACTGAGGCACTGAGGCACTGAGGCACTGAGGCACTGAGG 86730
Qy 374 a1e1Pro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 394
Db 86731 CACCATCTGTCT 86790

Qy 394 u1a1Pro1eupG1y1H1e1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 414
Db 86791 GGCACCTCTGGGAGCAAGAAACCAAGCACTGAGGCACTGAGGCACTGAGGCACTGAGG 86850
Qy 414 gG1y1eupG1a1a1y1Pro1eup1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1 433
Db 86851 GGGCTCTCAGGCGCCCTCTGCTCAGAGGCTCCACCTGATGAGGCGGAGCACTGAGG 86910
Qy 433 gG1y1eupPro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 453
Db 86911 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 86970
Qy 453 u1H1e1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 469
Db 86971 ACACATCACCACCAAGCACTGAGGCACTGAGGCACTGAGGCACTGAGGCACTGAGG 87019

RESULT 10
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US2003077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Alignment Scores:
Pred. No.: 4,49e-91 Length: 2379
Score: 1067.00 Matches: 257
Percent Similarity: 58.08% Conservative: 34
Best Local Similarity: 51.30% Mismatches: 100
Query Match: 42.70% Indels: 110
DB: 10 Gaps: 16

US-09-993-966-7 (1-470) x US-09-764-891-5902 (1-2379)
Qy 29 Trp1a1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 48
Db 82 TGGGCACTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141
Qy 49 Gly1Pro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 68
Db 142 TGGTACCCAGGCTTGGTAAAGCACTATTAAGGAGGAGGAGGAGGAGGAGGAGGAGG 198
Qy 69 Asp-Val1eup1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 88
Db 199 GAAATGACCTTAATAATGTTCAATAATTAAG-----TAATCATGACACGAAATGCA 252
Qy 88 a1eupPro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 108
Db 253 ACAAAATCAATATGATCTTAAGCTTAAGCACTTAAGCACTTAAGCACTTAAGCACTTA 312
Qy 108 lSerGlu1eupPro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 114
Db 313 ATCCAGCCCAACCAAGCCTGTTCCGGGCAACCACTTGAACCTCTCTGCTGTTTC 372
Qy 115 -Ser1y1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 134
Db 373 TTCAGCAAGAACCTGCTTCCCTTAAGTACAGCACTTACTGTTCTTTA----- 421
Qy 134 rArgGlu1eupPro1a1y1SerGlu1eupPro1a1y1ThSerg1nph1e1y1SerGlu1eupPro1a1y1 154
Db 421 ----- 421

```

QY 154 uAaplleThrSerleuLeuHisThrIleTyrgluValAlaAspSerSerValaenHis 174
Db 422 -----TCAGCTTGACATATGTGACCTTGCTCAATTCAC 459
QY 174 rProthSer----- 177
Db 460 TCCACAACTCCCTTCCTCCCTTCTCTCATATGTTAAATCTCATAGTACTTCT 519
QY 178 -----SerlysmetLeuArgValLysLeuThyValAlaProAsp----- 190
Db 520 GTATGTTTTAGCAATATACTTAGATT---CTTCTTGCCCCAGTACCTTAGATAGTT 576
QY 191 -----GlySerGlnSerLysArgSerValLeuVa 200
Db 577 TTCGGGCTCTGCTGCCATCTGTCTTCTGCGCTCTCTCGAG--AGTATTTT-- 631
QY 200 lAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLe 220
Db 632 -----GAAGACCTA-----CCAAAGAACCACTTCTCCA-----CT 663
QY 220 uArgSerTrpGluLys-----LysGlnArgAlaProLeuArg----- 232
Db 664 TAGCAGATGGGAAAGTTCCATTTTCAGAGAGTTTCTGCAAGGAGTGAAGGGGAGTCAG 723
QY 233 -----PheGlnGlyAspSerArgLeuGlnSerGlyCysTrpHisCysValaAspG 251
Db 724 GGCATTGGGGGTTGCTGCTGTTTGGAGAACTGGGGCG-----GGGCTGAATGC 774
QY 251 uAsnIleGluArgArgAsnHisTrpLeuAspLeuAlaGlyIleGluAsnTrpThrSerG 271
Db 775 TGGGGTATAGCGCAAGCCCAAGC-----ACAGTAGTGCTCATTAATATGCTGTGA 828
QY 271 nPheGlyPro-GlySerProSerValAlaGlnLysSerGluLeuProProArgThrSerA 291
Db 829 ATTGCTTCCAGAGCTCCCTTCCTGCGCCAGAACTGCAAGACTGCCCGCCAGCTTCA 888
QY 291 snProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysProG 311
Db 889 ATCCCATCTCAATCTCGCTCCCATAGCCGGAAGCCATCCATCCACACCGAAAGCCCC 948
QY 311 lngIValaAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGlu 331
Db 949 AAGGCGTGAACCGGCGCTCTTCCACTTCCTTGACACCCCAATGCCAAGSTCTCAAGGC 1008
QY 331 euGlnGlnArgLeuArgGlyThrGlnAspGlySerLysPheValArgSerPolya 351
Db 1009 TCCAGCAACGGCTCCGGGGCACCCAGAGCGGAGCAAGCACTTGTGTGAGTCCCCAAGG 1068
QY 351 laGlnGlyLysSerValGlyValGlyHisValAlaArgGlyValaArgAsnLysProPro 371
Db 1069 CCCAGGCGCAAGAGTGTGGGTGTGGGCCAGTGGCCAGAGGGGCAAGAAACAGCCCTTC 1128
QY 371 euGlyProAlaIleProAlaValSerProSerAlaHisIleuAlaIleSerProAlaLeu 391
Db 1129 TGGGACCGCGCATCTCGCGGTGTCCCTCGGCCACTGTGCTGCCAGCCCGGCTCC 1188
QY 391 euProSerLeuAlaProLeuGlyHisLysLysHisLysHisArgAlaLysGluSerGln 411
Db 1189 TCCCTCTCTAGCCCTCCGGGCAAGAAAGCAAGACCGGACCAAGAGAGCCGCGC 1248
QY 411 lngIYcysArgGlyLeuGlnAlaProLeuAlaSerGlyLysProValLeuGlyArgGlu 431
Db 1249 AGGGCTGCGGGGCTGAGGACCACTGGCTCAAGGCGCTGTCTTGGGGGAGGAGC 1308
QY 431 lAsLeuArgGluLeuProAlaLeuValValTyrgLysGlnAlaGlyGlnProValGlna 451
Db 1309 ACCTGCGGAGAGCTCGCGCTTGGTGTGTATGAGCAAGCCGGGAGCGGTCCAGA 1368
QY 451 rghIsgIuHisHisHisHisIsgIuHisHisHisHisArgTrpHisHisPheTyrglnThr 470
Db 1369 GACATGAGCAACACACCATGAACTCAACACATTAACACACACTTTCACAGACA 1427

```

```

RESULT 11
US-09-764-891-5904
; Sequence 5904, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5904
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5904

Alignment Scores:
Pred. No.: 4,49e-91 Length: 2379
Score: 1067.00 Matches: 257
Percent Similarity: 58.08% Conservative: 34
Best Local Similarity: 51.30% Mismatches: 100
Query Match: 42.70% Indels: 110
DB: Gaps: 16

US-09-993-966-7 (1-470) x US-09-764-891-5904 (1-2379)

QY 29 TrpAlaArgLysGlyIleGlnGluTrpIleGlyArgGlnArgCysProGlyGlyValSer 48
Db 82 TGGGCGATGAGGAGGCGCTGCGTGAACGGCCAGCGCGCGCGGTGTGTTCAC 141
QY 49 GLyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSerThrArgGluLeuValGly 68
Db 142 TGCTAACCCAGGCTCGCTTAAGCAACTATTATGGACAGTCMAAGA---CTTCTTGA 198
QY 69 Asp-ValLeuArgAspThrLeuSerGlnGlnGluAlaAspPheArgLeuGluValAla 88
Db 199 GAAGTACACTTAATAATGTTCATAAATTAAAG-----TAATACATGACGAAATCGA 252
QY 88 aLeuProGluGluThrAspGlyLeuGlySerGlyAspGlyLysLysMetGluArgVa 108
Db 253 ACAAAATCCATATGATCTTAAGCTTAGCACCCATGACAAACAAACAGTTTGTGCCCC 312
QY 108 lSerGluProCys-----ProGly----- 114
Db 313 ATCCACAGCCCAACCCAGCCCTGTTCGCGGGCAACACTTGACCTCTCGGCTGTTC 372
QY 115 -SerLyLeuGlnLeuLysPheGlnGluLeuGlnCysAspValSerMetGluAspSe 134
Db 373 TTCAGCAAGACCTCGGTTTCCCTAAGTACAGACTACGTGTCTTA----- 421
QY 134 rArgGlnGluTrpThrPheThrLeuTyraAspPheAspAsnGlyLysValThrArgG 154
Db 421 ----- 421
QY 154 uAaplleThrSerleuLeuHisThrIleTyrgluValAlaAspSerSerValaenHis 174
Db 422 -----TCAGCTTGACATATGTGACCTTGCTCAATTCAC 459
QY 174 rProthSer----- 177
Db 460 TCCACAACTCCCTTCCTCCCTTCTCTCATATGTTAAATCTCATAGTACTTCT 519
QY 178 -----SerlysmetLeuArgValLysLeuThyValAlaProAsp----- 190
Db 520 GTATGTTTTAGCAATATACTTAGATT---CTTCTTGCCCCAGTACCTTAGATAGTT 576
QY 191 -----GlySerGlnSerLysArgSerValLeuVa 200
Db 577 TTCGGGCTCTGCTGCCATCTGTCTTCTGCGCTCTCTCGAG--AGTATTTT-- 631

```

```

QY 200 LaenGlnAlaAspLeuGlnSerAlaArgProAlaGluThrIleProThrIleAspLe 220
Db 632 -----GAAAGACCTA-----CCAAAGAACGATTCTTCCA-----CT 663
QY 220 uArgSerTrpGluIle-----LysGlnArgAlaProLeuArg----- 232
Db 664 TAGAGATGGGAAAGTTCATTCAGAGAGTTTCTGCAAGGACAGAGGAGGAGCAG 723
QY 233 ----PheGlnGlyAspSerArgLeuGlnInserGlyCysTrpIleCysValAspG 251
Db 724 GGCAATTGGGGGTGCTGCTGTTTGGAGAACTGGGGCGC-----GGGGTATGTC 774
QY 251 uAenIleGluArgArgAsnIleTrpLeuAspLeuAlaGlyIleGluAsnTrpIleSerG 271
Db 775 TGGGGTATGAGCGCAAGCCCAAGCAC-----ACAGTAGGTGCTCAATTAATGTCGTTGA 828
QY 271 nPheGlyPro-GlySerProSerValAlaGlnLysSerGluLeuProProArgThrSera 291
Db 829 ATTGGTTCCTAGAGCTCCCTTCCTGGCCCAAGAGTCAGAACTGCCCCCGCACCTTCA 888
QY 291 snProThrArgSerArgSerIleGluProGluAlaIleIleIleProIleArgIleArgProG 311
Db 889 ATCCCATTCATCTCGCTCCCATAGCCGGAAGCCATCCATCCACACCGAAAGCCCC 948
QY 311 InGlyValAspProAlaSerPheIleAspLeuAspThrProIleAlaLysValSerGlu 331
Db 949 AAGGCGTGAGCCGGCTCTCCCTTCACATTCCTGACCCCAATGCCAAGGTTCACAGC 1008
QY 331 euGlnGlnArgLeuArgGlyThrGlnAspGlySerLysIlePheValArgSerProLysA 351
Db 1009 TCCAGCAACGGCTCCGGGGGACCCAGAGACGGAGCAACGACTTTGTAGGTCCCCCAAG 1068
QY 351 laGlnGlyLysSerValGlyValGlyIleValAlaArgGlyValArgAsnLysProIle 371
Db 1069 CCGAGGCAAGAGTGTGGTGGGCGCACGTCGGCCAGAGGGGCAAGAACAGCCCTCC 1128
QY 371 euGlyProAlaIleProAlaValSerProSerAlaIleLeuAlaIleSerProAlaLeu 391
Db 1129 TGGAGCCCGGATCTCCGCGGTGCTCCCTCCCGCCACTGGCGCCAGCCCGGCTCC 1188
QY 391 euProSerLeuAlaProLeuGlyIleLysLysIleLysIleAspAlaLysGlySerGlnG 411
Db 1189 TCCCTCTCCAGCCCTCCCTGGGACAAAGACAAAGACAAAGCCCAAGAGAGCCAGC 1248
QY 411 InGlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyIleProValLeuGlyArgGlu 431
Db 1249 AGGGCTGCCGGGGCTGACAGGACCACTGAGTGGCCCTGCTCGGGGGGGAGC 1308
QY 431 IseuArgGluLeuProAlaLeuValValTrpGluSerGlnAlaGlyGlnProValGlnA 451
Db 1309 ACCTGCGGAGGCTGCCCGCTGTGTGTATGAGACCGAGCCGAGCCGAGTCCAGA 1368
QY 451 rghIleGluIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 470
Db 1369 GACATAGACCAACCAACCAATGACATCAACCACTTACCACTTACCAAGACA 1427

RESULT 12
US-09-993-966-21
; Sequence 21, Application US/09993966
; Publication No. US2003018623A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NMD PROTEIN,
; NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/028073
; CURRENT APPLICATION NUMBER: US/09/993,966
; PRIOR FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01

```

```

; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 21
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-21

Alignment Scores:
Pred. No.: 1,59e-90 Length: 590
Score: 1053.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.14% Indels: 0
DB: Gaps: 0

US-09-993-966-7 (1-470) x US-09-993-966-21 (1-590)

QY 276 SerProSerValAlaGlnLysSerGluLeuProProArgThrSeraProThrArgSer 295
Db 3 TCCCTTCCTGAGCCGGAAGTCAGAACTGCCCCCGCACCTTCATCCACTCGATCT 62
QY 296 ArgSerIleGluProGluAlaIleIleIleProIleArgIleArgProGlnIleValAspPro 315
Db 63 CGCTCCATAGCCGGAAGCCATCCATCCACACCGAAAGCCCAAGGCGTGAGCCG 122
QY 316 AlaSerPheIleAspLeuAspThrProIleAlaLysValSerGluLeuGlnArgLeu 335
Db 123 GCCTCTTCACATTCCTTGACACCCCAATGCCAAGGTTCAGAGTCCAGCAACGGCTC 182
QY 336 ArgGlyThrGlnAspGlySerLysIlePheValArgSerProLysValGlnGlyLysSer 355
Db 183 CCGGGCACCCAGAGACGGAGCAACGACTTTGTAGGTCCCCCAAGGCCAGGCGCAAGAGT 242
QY 356 ValGlyValGlyIleValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIle 375
Db 243 GTGGGTGGGCGCACGTCGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGCCGCCATC 302
QY 376 ProAlaValSerProSerAlaIleLeuAlaIleSerProAlaLeuLeuProSerLeuAla 395
Db 303 CTGCGGTGTCCTCCCTCCCGCCACTGGGTGCCAGCCCGGCTCCCTCCCTCGTAGCC 362
QY 396 ProLeuGlyIleLysLysIleLysIleAspAlaLysGlySerGlnGlnGlyCysArgGly 415
Db 363 CCCCCTGGGCAAGAGCAACAGCAACCGACCAAGAGAGCCAGCAGGCGCTCCCGGGC 422
QY 416 LeuGlnAlaProLeuAlaSerGlyIleProValLeuGlyArgGluIleLeuArgGluLeu 435
Db 423 CTGCAAGCACCACTGGCTCAGGTGGCCCTGCTCGGGGGAGACCTTGGCGGAGCTG 482
QY 436 ProAlaLeuValValTrpGluSerGlnAlaGlyGlnProValGlnArgIleGluIleIle 455
Db 483 CCGGCTGTGTGTATGAGACCGAGCCGAGCCGAGTCCAGAGACATGAGCAACAC 512
QY 456 HisIleHisGluIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 470
Db 543 CACCACATGATGATCAACCACTTACCACTTACCACTTACCAAGACA 587

RESULT 13
US-09-764-868-152
; Sequence 152, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 152
; LENGTH: 1966

```


Pred. No.: 1 57e-64 Length: 1959
 Score: 786.00 Matches: 215
 Percent Similarity: 47.24% Conservative: 50
 Best Local Similarity: 38.32% Mismatches: 152
 Query Match: 31.45% Indels: 145
 DB: 18 Gaps: 20

US-09-993-966-7 (1-470) x US-10-723-860-5220 (1-1959)

```

QY 1 MetGlyLeuLeuHisSerLysProAlaAlaValCys---LysArgArgLysSerProGlu 19
DB 98 ATGGGGAAATCGACGTGGAAGACAGCCGCCGCCGCAAGCGAGAGAGAGAGAGAGAGAGAG 157
QY 20 GlyAspSerLeuAlaValSerAla---AlaTrpAlaArgLysGlyIleGluTrpIle 38
DB 158 GGGGACAGCTTCTGTCGGCTCCGCTACCGCGCGCGCGCAAGAGCGCGGAGAGAGAGAG 217
QY 39 GlyArg-----GlnArgCysProGlyGly-----ValSerGlyProArgGln 52
DB 218 CGGCGCGCGCGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 277
QY 53 LeuArgLeuAlaGlyThrIleGlyArgSerThrArgLysLeuValGlyAspValLeuArg 72
DB 277 ----- 277
QY 73 AspThrLeuSerGluGluGluGluAspPheArgLeuGluValAlaLeuProProGlu 92
DB 278 -----GAGGACCAAGTGTCTCCCTACAGGTGTGACATCCCGCTGAG 316
QY 93 LysThrAspGly-----LeuGlySerGlyAsp-----GluLysLys 104
DB 317 AAGGCTGAGGAGCGCGAGACCGCGGCAAACTCTCAAGAGATGAGAGAGAGAGAGAGAG 376
QY 105 MetGluArgValSerGluProCysProGlySerLysGlnLeuLysPheGluGluLeu 124
DB 377 GCAACCGGAGGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 433
QY 125 GlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPheThrLeuTyrAsp 144
DB 434 CAGTGGAGTGTCTCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
QY 145 PheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisIleTyr 164
DB 494 TTGGACAACTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 553
QY 165 GluValValAspSerValAsnHisSerProThrSerSerLysMetLeuArgValLys 184
DB 554 GAGGTGTGATGCTCGTCAACCACTCTCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 613
QY 185 LeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuValAsnGlnAlaAsp 204
DB 614 CTAAACGCTACGCGCTGAGCCCTCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 673
QY 205 LeuGlnSerLysArgProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGlu 224
DB 674 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 733
QY 225 LysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGluGlnSerGlyCys 244
DB 734 AGGAGGTGTCTCTCCACAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 793
QY 245 TyrHisHisCysValAspGluAsnIleGluArgArgAsnHisGlyLeuAspLeuAlaGly 264
DB 794 GGGCCCTTACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 853
QY 265 IleGluAsnThrThrSerGlnPheGlyProGlySerProSerValAlaGlnLysSerGlu 284
DB 854 ATTGAAGAACTACAGTCAAGATTCGCGCTGAGTCTCAAGTGTGCGAGAGAGAGAGAG 913
QY 285 LeuPro----- 286
DB 914 GCTCCAGAGACACAGAGTGGGAGCAAGGCTAAGAGAGTGGCGCTTTGACAGAGAGCTGTGG 973
  
```

```

QY 286 ----- 286
DB 974 AGCCAGGACAGGTACCCACAGTGGCCAGGCCCTTCCCTTACAGGCTGTGGCGGTGA 1033
QY 286 ----- 286
DB 1034 CTGACACTTGGCTACAGACTGAGCTCAGAGTCCCTCTGTGCAAGAGAGAGAGAGAGAG 1093
QY 287 -----ProArgThrSerAsnProThrArgSerArgSerHisGluProGluAla 302
DB 1094 CCCAGGAGAGAGAGCTGTGACCTCCAGAGCCCGG-----TCCGCTCCAGAGAGAGATACA 1149
QY 303 IleHisIleProHisArgGlySerProGln-----GlyValAspProAlaSerPhe 318
DB 1150 CATGCCCTACACACACCGACAGTCAAGAGTGTGTGAGACAGTCGTGCGAGCTCG--- 1206
QY 319 HisPheLeuAspThrProIleAlaLysValSerGluLeuGlnIleArgLeuArgGlyThr 338
DB 1207 -----GAGCTGTGCTCCGAGGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
QY 339 GlnAspGlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyVal 358
DB 1255 GAG-----AAGAGTTCCTCAAGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
QY 359 GlyHisValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaVal 378
DB 1306 NNNNNNNNNNNNNNNNTCCGGGAAA-----GCTTCAGACTACTACCTGCGCGCTC 1356
QY 379 SerPro-----SerAlaHisLeuAlaAlaSerProAlaLeuLeuPro 392
DB 1357 CTGCGCGCCAGGAGCCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407
QY 393 SerLeuAlaProLeuGlnHisLysLysLysHisArgAlaLysGluSerGlnGlnGly 412
DB 1408 -----CCACTTACGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1452
QY 413 CysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGluHis--- 431
DB 1453 CACTCGCACTCAAGAGCCCAACAGCTCAGCTGACAGTGTG-----GAGCAGAGAG 1503
QY 432 ---LeuArgLysLeuProAlaLeuValAlaTyrGluSerGlnAlaGlyGlnProValGln 450
DB 1504 GTGTGCGGAGACTGCGCGCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563
QY 451 ArgHisGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 467
DB 1564 CGGACGAGAGACACACACACACACACACACACACACACACACACACACACACACACACTTC 1623
QY 468 Tyr 468
DB 1624 CAC 1626
  
```

RESULT 15
 US-09-918-995-8686
 ; Sequence 8686, Application US/0918995
 ; Publication No. US2003073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyeq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918, 995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235, 076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: PatcSBQ for Windows Version 3.0
 ; SEQ ID NO 8686
 ; LENGTH: 427
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature

LOCATION: (1)...(427)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8686

Alignment Scores:

Pred. No.:	6,37e-61	Length:	427
Score:	739.00	Matches:	140
Percent Similarity:	99.29%	Conservative:	0
Best Local Similarity:	99.29%	Mismatches:	1
Query Match:	29.57%	Indels:	0
DB:	10	Gaps:	0

US-09-993-966-7 (1-470) x US-09-918-995-8686 (1-427)

QY	285	LeuProProArgThrSerAnProThrArgSerArgSerHisgluProgluAlaIleHis	304
DB	4	CTGCCCCCGCACCTCCAAATCCCACTCGCTCCATGAGCGGAAGCCATCCAC	63
QY	305	ILEProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeuAspThrPro	324
DB	64	ATCCCAACCCGAAGCCCAAGGGGTGACCCGGCTCTTCCATTCTTGACACCCCA	123
QY	325	ILEAlaLysValSerGlnLeuGlnArgLeuArgGlyThrGlnAspGlySerLysHis	344
DB	124	ATGCGCAAGGTCTCAGAGCTCCAGCAAGGCTCCGGGGCAACCGAGCGGAGCAAGCAC	183
QY	345	PheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHisValAlaArgGly	364
DB	184	TTTGTGAGTCCCGCCAGGCGCCAGGGCAGAGTGTGGGTGGGCCACGTGGCCAGAGGG	243
QY	365	AlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSerAlaHisLeu	384
DB	244	GCAAGAAACAAAGCCCTCTGGGACCCGCCATCCCTGGGTGTCCTCCGCCCACTG	303
QY	385	AlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysLysHis	404
DB	304	GCTGCCAGCCCGGGCTCTCCCTCCCTAGCCCCCTCGGGCAACAAGACAAAGCAC	363
QY	405	ArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyGly	424
DB	364	CGAGCCAAAGAGAGCAAGGGGTGCGGGGCTTGCAAGCACCACTGGCTCAAGTGGC	423
QY	425	Pro	425
DB	424	CCT	426

Search completed: December 30, 2004, 09:35:19
Job time : 5904 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 29, 2004, 21:21:19 ; Search time 848 Seconds

(without alignments)

2909.467 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 2499

Sequence: 1 MGKLSKRAVCKRRESREG.....RHHNNHHNNHHNNHHYQT 470

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool_P/US0993966/runat.29122004.130757.8986/app.query.fasta_1.647
-DB=N Geneseq.23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCORB=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US0993966_@CGN.1.1.885_@runat.29122004.130757.8986 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_23Sep04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001a:*\n5: geneseqn2001b:*\n6: geneseqn2002a:*\n7: geneseqn2002b:*\n8: geneseqn2003a:*\n9: geneseqn2003b:*\n10: geneseqn2003c:*\n11: geneseqn2003d:*\n12: geneseqn2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2499	100.0	1859	10	AAL51511
2	2493	99.8	1438	3	AAAG63925
3	2192.5	87.7	1416	4	AAH28343
4	2174.5	87.0	1731	3	AAAG63924
5	2171.5	86.9	1401	10	AAL51512
6	1067	42.7	2379	4	AAL03214

7	1067	42.7	2379	4	AAL03216	AAL03216 Human rep
8	1053	42.1	590	10	AAE51525	AAE51525 Human Nkd
9	842	33.7	1743	5	AA878752	AA878752 DNA encod
10	842	33.7	1743	5	AA892081	AA892081 DNA encod
11	842	33.7	1743	10	ADC31966	ADC31966 Human nov
12	831.5	33.3	1966	4	AA827117	AA827117 CDNA enc
13	831.5	33.3	1966	10	AD893295	AD893295 Human CDN
14	798.5	32.0	1561	5	AA892082	AA892082 DNA encod
15	786	31.5	1959	12	ADQ22400	ADQ22400 Human sof
16	783	31.2	1285	3	AAAG63926	AAAG63926 DNA encod
17	780.5	31.3	1885	10	ADFE82455	ADFE82455 Leukaemia
18	739	29.6	427	9	ACH21474	ACH21474 Human adu
19	583.5	23.3	1307	3	AAAG63927	AAAG63927 DNA encod
20	378.5	15.1	1817	6	ABK34945	ABK34945 Human CDN
21	271	10.8	639	8	ACD05637	ACD05637 cDNA enco
22	246	9.8	1950	6	ABQ73820	ABQ73820 Human col
23	239	9.6	598	12	ACH74795	ACH74795 Human gen
24	238	9.5	148	10	AAU51522	AAU51522 Human Nkd
25	237	9.5	128	10	AAU51524	AAU51524 Human Nkd
26	233	9.3	133	12	ACH88495	ACH88495 Human gen
27	233	9.3	134	10	AAU51518	AAU51518 Human Nkd
28	215.5	8.6	2691	4	ABL24299	ABL24299 Drosophill
29	215.5	8.6	4954	3	AAAG63923	AAAG63923 DNA encod
30	209	8.4	2723	4	AAK75944	AAK75944 Human imm
31	186	7.4	107	10	AAU51520	AAU51520 Human Nkd
32	182	7.3	3486	4	ABL27205	ABL27205 Drosophill
33	175	7.0	96	10	AAU51521	AAU51521 Human Nkd
34	172	6.9	6749	12	ADQ24048	ADQ24048 Human sof
35	170.5	6.8	2740	3	AACT5288	AACT5288 Human ORF
36	164	6.6	8104	4	ABL24298	ABL24298 Drosophill
37	164	6.6	51855	10	ACR04818	ACR04818 Melithiaz
38	161.5	6.5	554	12	ACH74420	ACH74420 Human gen
39	159	6.4	9807	4	ABL06949	ABL06949 Drosophill
40	159	6.4	23407	4	ABL06948	ABL06948 Drosophill
41	158.5	6.3	2776	4	ABL07259	ABL07259 Drosophill
42	156.5	6.3	1633	10	ADCG7022	ADCG7022 Human GPC
43	156.5	6.3	2981	5	AA870530	AA870530 DNA encod
44	156.5	6.3	2981	10	AD809862	AD809862 Novel DNA
45	156.5	6.3	6714	4	ABL16321	ABL16321 Drosophill

ALIGNMENTS

RESULT 1

AAL51511 standard; DNA; 1859 BP.

XX	AC	AAL51511;	
XX	DT	24-APR-2003 (first entry)	
XX	DE	Human Nkd (hNkd) protein coding sequence.	
XX	XX	Human, gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;	
XX	KM	colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.	
XX	OS	Homo sapiens.	
XX	FT	Key	Location/Qualifiers
XX	FT	CDS	224..1636
XX	FT		/tag= a
XX	FT		/product= "Human Nkd protein"
XX	PD	21-NOV-2002.	
XX	XX	27-NOV-2001; 2001WO-US044092.	
XX	PF	27-NOV-2000; 2000US-0252884P.	
XX	PR	16-MAY-2001; 2001US-0291109P.	
XX	PR	01-OCT-2001; 2001US-0325571P.	

PA (CHIR) CHIRON CORP.
 XX Rohan M, Chan V, Yan D;
 XX WPI; 2003-129303/12.
 DR P-PSDB; AAO16389.
 XX
 PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
 PT useful for treating cancer involving aberrant Wnt signaling, e.g. colon
 PT cancer, head and neck cancer, ovarian cancer, or breast cancer.
 XX
 PS Claim 4; Fig 3; 99pp; English.
 XX
 CC The invention comprises the amino acid and coding sequence of the human
 CC Nkd protein (hNkd) - a regulator of Wnt signaling pathways. The hNkd DNA
 CC and protein sequences are useful for the treatment of cancer involving
 CC aberrant Wnt signaling (e.g. colon cancer, head and neck cancer, ovarian
 CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
 CC protein of the invention
 XX
 SQ Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 2.87e-147 Length: 1859
 Score: 2499.00 Matches: 470
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-993-966-7 (1-470) x AML5151 (1-1859)
 QY 1 MercGlylsrleuHisSerlySProlaAlaValCylsYsarGaTgGluSerProGluGly 20
 Db ATGGGGAAACTTCACTCCAAAGCCGGCCGCTGTGCAAGCGAGAGAGCCCGAAAGGT 283
 QY 21 AspSerPheAlaValSerAlaAlaTPrAlaRglYsglyIleGluGluTPrIleGlyArg 40
 Db GACAGCTTCGCGGAGGCTGCTGGGCTCGAAAGGCGCATCGAGTGGATTCGGAGGA 343
 QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
 Db CAGGCGTCCCGGGCGGGTGTCTGGGACCCCGACAGCTGGCGTGGGGGACCATAGGC 403
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
 Db CGAAGCACCCGGGAGCTCTGGGCGACGTGTGAGAGACACGCTCAGCGAAGAGAGAG 463
 QY 81 AspAspPheArgGluGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
 Db GACGACTTTCGGCTGGAGAGTGGCCCTGCTCTGAGAGACTGACGGGCTGGGCAAGCGA 523
 QY 101 AspGluLysLysMetGluArgValSerGluProCysProGlySerLysGluLeuLys 120
 Db GATGAGAAAGAGATGAGAGAGTGAAGGACCTCGCCAGGCTCCAGAAACAGACTGAG 583
 QY 121 PheGluGluLeuGluCysAspValSerMetGluGluAspSerArgGluGluTPrThrPhe 140
 Db TTTGAAAGGCTCCAGTGGAGCGTGTCCATGAGAGAGACAGCCCGCAGAGTGAACCTTC 643
 QY 141 ThrLeuTyrAspPheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeu 160
 Db ACCCTGTATGACTTTGACACACACGGAAGGTACCCGAGAGGACATCACAGGCTTGCTG 703
 QY 161 HisThrIleTyrGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
 Db CACACCATCTATGAGGTGTGACTCCTCTCAACACACTCCCAACATTCACGAGAAGTG 763
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGluSerLysArgSerValLeuVal 200
 Db CTGGGGGTAAAGCTCACCGTGGCCCGCATGTGCGACGACCAAGAGAGCGTCTTGTTC 823
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgLagIuThrLysProThrGluAspLeu 220

Db 824 AATCAGCTGACTTCGACAGCGCAGAGCCCGGAGAGACCAAGCCCATGAGGACTG 883
 QY 221 ArgSerTrpGluLysLysGlnAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 Db CGGAGCTGGGAGAAAGACAGCAGAGCCCGCTCAGGTTCACGGGTGACACCGCTGGAG 943
 QY 241 GluSerGlyCysTyrThrShiShiCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
 Db CAGTCTGGCTGTACCACTCCATTCGTATGATGAGAACATCGAAGAGAAACCACTACTTA 1003
 QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 Db GATCTCCCGCGGATAGAAACTACAGTCCCATTTGGGCTGCTCCCTCCGTGGCC 1063
 QY 281 GluLysSerGluLeuProProAlaGlyThrSerAsnProThrArgSerArgSerHisGluPro 300
 Db CAGAACTCAGAACTGCCCCCCCCCACCCTCCAAATCCCATCTGATCTGCTCCATGAGCCG 1123
 QY 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
 Db GAAAGCATCCACATCCACACCGAAGGCCCAAGGCGTGAACCGGCTCTTCACTTC 1183
 QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyTyrGlnAsp 340
 Db CTTGACACCCCATTCGCGCAAGGTCTCAGAGCTCCAGAACAGGCTCCGGGACACCAAGAC 1243
 QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
 Db GGGAGCAGAGCATTTGTGAGGTCCCCCAAGGCCAGGCAAGAGTGTGGGTGGGCGAC 1303
 QY 361 ValAlaArgGlyValAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 Db GTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAGCCGCCATCTCGGGGTGTCCCCC 1363
 QY 381 SerAlaHisLeuAlaLysSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 Db TCGGCCACTGTGCTGCACCGCGGCTCTCCCTCCCTTAAGCCCCCTCGGGACAAAG 1423
 QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLysLeuGlnAlaProLeu 420
 Db AAGCACAAAGCACCGAGCCAGAGAGCCAGAGGCTGCCGGGCTCGAGGACCACTG 1483
 QY 421 AlaSerGlyLysProValLeuGluArgGluHisGluArgGluLeuProAlaLeuValAla 440
 Db GCTTCAGGTGGCTGTCTGTGGGGGAGACACTCTGCGGAGCTGCCGCTTGCTGTG 1543
 QY 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisShiShiShiShiGluHis 460
 Db TATGAGAGCCAGGCGCGGCGAGCCGGTCCAGAGACATGACACCAACCATGAACAT 1603
 QY 461 HisHisIleTyrHisShiShiPheTyrGlnThr 470
 Db CACCACCATTAACCACTTCTACAGACA 1633
 Db
 RESULT 2
 AAA63925
 ID AAA63925 standard; DNA, 1438 BP.
 AC AAA63925;
 XX
 XX 04-DEC-2000 (first entry)
 XX
 XX DNA encoding protein related to Drosophila naked cuticle polypeptide.
 XX Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signaling;
 KW gene therapy; Nkd defect; cancer; ss.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 8..1420
 FT CDS
 FT /*tag= a

FT /product= "naked cuticle polypeptide"
 XX WO200049034-A1.
 PN 24-AUG-2000.
 PD 17-FEB-2000; 2000MO-US004188.
 XX 17-FEB-1999; 99US-0120646P.
 PR (STRD) UNIV LEIAND STANFORD JUNIOR.
 PA Scott M, Zeng W, Wharton K;
 XX WPI; 2000-571967/53.
 DR P-PSDB; AAB08216.
 XX An isolated nucleic acid molecule useful for analyzing (genetic
 PT predileposition to) a disease state and for therapeutic purposes e.g.
 PT treatment of cancer comprises a sequence encoding a naked cuticle
 PT protein.
 PS Claim 3; Page 46-48; 58pp; English.
 XX The present sequence encodes a protein related to the Drosophila Nkd
 CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
 CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide
 CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the
 CC regulation of Wnt signal potency, duration or distribution. The Nkd
 CC polynucleotide can be used for identifying homologous or related
 CC proteins, to modulate the expression or function of Nkd polypeptides, and
 CC in studying associated physiological pathways. Nkd polynucleotides can
 CC also be used in gene therapy to treat disorders associated with Nkd
 CC defects. They may also be used for therapeutic purposes e.g. treatment of
 CC cancer
 XX
 SO Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;

 Alignment Scores:
 Pred. No.: 5.08e-147 Length: 1438
 Score: 2493.00 Matches: 469
 Percent Similarity: 99.79% Conservative: 0
 Best Local Similarity: 99.79% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 3 Gaps: 0

 US-09-993-966-7 (1-470) x AAA63925 (1-1438)
 QY 1 MercGlyLylLeuHissErlYpTroAlaValCylsYsArGArgGluSerProGluGly 20
 DB 8 ATGGGAAATCTTCACTCCAAAGCCGCGCGCTGGCAAGCGAGGAGACCCGGAAGGT 67
 QY 21 AspSerPheAlaValSerAlaAlaTPAlaArgLysGlyTleGluGluTPIleGlyArg 40
 DB 68 GAGAGCTTCCTCCGAGCGCTGCTGGCTCGAAAGGCATCGAGAGTGAATCGGAGA 127
 QY 41 GlnArgCyseProGlyLysGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
 DB 128 CACGCGTGGCGCGGCTGTCTGGGACCCGACAGCTGGGCTGGGCGACCATGGC 187
 QY 61 ArgSerThrArgGluLeuValIGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
 DB 188 CGAAGACCCCGAGCGCTGCGGCGAGAGTGTGAGACACGCTCAAGCGAGAGAGAG 247
 QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
 DB 248 GAGACCTTTCGCTGGAAGTGGCCCTCTCTCGAAGAGCTAGCGGCTGGGCGAGCGA 307
 QY 101 AspGluLysLysMetGluArgValSerGluProCyseProGlySerLysGlnLeuLys 120
 DB 308 GATGAGAAAGAGATGGAGAGAGTGAAGACCCGCGCCAGGCTCCAAAGAGAGCTGAAG 367
 QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140

DB 368 TTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACGCGGCGAGGTGACCTTCC 427
 QY 141 ThrLeuTyraSPheAspAenAenGlyLysValThrArgGluAspIleThrSerLeuLeu 160
 DB 428 ACCCTGTATGACTTTGACCAACAAAGGAGTCAACCGAGAGAGACATCACACGCTTCTG 487
 QY 161 HisThrIleTyrgluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180
 DB 488 CACACCATATAGAGGTGTGAGACTCTCTGTCAACACCTCCCAATCATCAGAGATG 547
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200
 DB 548 CTGCGGGTAAAGCTCACCGTGGCCCCGATGGAGCAGCGAGCAAGAGAGCGTCTTGTG 607
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
 DB 608 AATCAGGCTGACCTGGAGAGCGGAGGCCCCGAGAGAGACCAAGCCACTGAGAGACTG 667
 QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 DB 668 CGGAGCTGGGAGAAAGACAGCAGCGCCCGCTCAGGTTCCAGGGGTGACAGCCGCTGGAG 727
 QY 241 GlnSerGlyCyseThrHisGlyAspValAspGluAsnIleGluArgArgAsnHisTyrglu 260
 DB 728 CAGTCTGGCTGTACCACTTCGTATGATGAGAACATCCAGAGAGAAACCACTACTTA 787
 QY 261 AspLeuAlaGlyIleGluAsnTyrglnSerGlnPheGlyProGlySerProSerValAla 280
 DB 788 GATCTCCCGGATAGAAACCTACAGTCCCAATTTGGGCTGCTCCCTTCCTGCTGGCC 847
 QY 281 GlnLysSerGluLeuProProArgTyrglnSerAsnProThrArgSerArgSerHisGluPro 300
 DB 848 CAGAGTCAGAAATGCGCCCCCCCCACCTTCATATCCACTGATCTGCTCCCTGAGCCG 907
 QY 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
 DB 908 GAAAGCATCCACATCCCAACCGAAGAGCCCAAGGCGTGAACCCGCGCTCTCCACTTC 967
 QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
 DB 968 CTTGACACCCCAATCCCAAGATCTCAGAGCTCAGACAGCGCTCCGGGCGACTCAGAGAC 1027
 QY 341 GlySerLysHisAspHeValArgSerProLysValGlnGlyLysSerValGlyValGlyHis 360
 DB 1028 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCCAAGGCAAGAGTGGGTGGGCGCAC 1087
 QY 361 ValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValaSerPro 380
 DB 1088 GTGGCCAGAGGGGCAAGAAACAAAGCCCCCTCTGGAGCCCGCATCCCTGGGTGTCCCCC 1147
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 DB 1148 TCGGCCACCTGGCTGCCAGCCCGGCTCTCTCCCTCTTACCCCCCTCGGGCACAAG 1207
 QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 DB 1208 AAGCACAAGACACGAGACCAAGAGAGACAGAGGAGCTGCGGGGCTGCGAGGACCACTG 1267
 QY 421 AlaSerGlyLysProValLeuGlyArgGlnHisLeuArgGluLeuProAlaLeuVal 440
 DB 1268 GCTTCAGGTGGCTCTCTCTGGGGCGAGAGACCTCGGAGGCTGGCCGCTTGTGGTG 1327
 QY 441 TyrglnSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisGluHis 460
 DB 1328 TATGAGAGCAGGCGCGGCGAGCGGCTCCAGAGACATGAGACCAACCACTATGAACAT 1387
 QY 461 HisHisHisTyrglnHisHisPheTyrglnThr 470
 DB 1388 CACCACCATTAACACCACTTCTACAGACA 1417
 RESULT 3
 AAH28343

ID AAH28343 standard; DNA; 1416 BP.
 XX
 AC AAH28343;
 XX
 DT 05-SEP-2001 (first entry)
 XX
 DE Nucleotide sequence of a murine mnkd protein.
 XX
 KM Dishevelled-associated protein 1A; DAP1A; mnkd; dishevelled protein;
 KW mt signalling; colon cancer; cancer; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1416
 FT /*tag= a
 FT /product= "mnkd"
 XX
 PN MO200144279-A2.
 XX
 PD 21-JUN-2001.
 XX
 PF 05-DEC-2000; 2000MO-US032986.
 XX
 PR 17-DEC-1999; 99US-0172434P.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Yan D, Williams LT;
 XX
 DR WPI; 2001-425440/45.
 DR P-PSDB; AAB84645.
 XX
 PT Novel mammalian dishevelled-associated proteins, mnkd and DAP1A, useful
 PT for inhibiting Wnt signaling in mammalian cells, and thus for treating
 PT colon cancer.
 XX
 PS Claim 2; Fig 1; 67pp; English.
 XX
 CC The present sequence encodes a murine mnkd protein. The specification
 CC describes mnkd and dishevelled-associated protein 1A (DAP1A) proteins,
 CC that interact with mammalian dishevelled protein. mnkd is useful for
 CC inhibiting Wnt signaling in mammalian cells, and thus for treating colon
 CC cancer. mnkd is also useful for activating the JNK pathway. mnkd and
 CC DAP1A are also useful for screening drugs that are useful for treating
 CC cancer.
 CC
 XX
 SQ Sequence 1416 BP; 339 A; 434 C; 419 G; 224 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 3,199-128 Length: 1416
 Score: 2192.50 Matches: 410
 Percent Similarity: 91.06% Conservative: 18
 Best Local Similarity: 87.23% Mismatches: 41
 Query Match: 87.74% Indels: 1
 DB: 4 Gaps: 1
 US-09-993-966-7 (1-470) x AAH28343 (1-1416)
 QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysIlySarGArgGluSerProGluGly 20
 Db 1 ATGGGGAACCTTCACTCGAAGCCGGCCGCTGTGCAAGCCAGAGAGCCCGGAAGT 60
 QY 21 AspSerPheAlaValSerAlaAlaTrpAlaArgLysGlyIleGluGluTrpIleGlyArg 40
 Db 61 GACAGCTTTGCTGTAAAGCGCTGTGGGCAAGAGCATCGAGAGTGGATCGGAGG 120
 QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
 Db 121 CAGGCGTGTCCAGGACGCGTTCAGAGACCCGTCAGAGATTGGCAGGCACTGTGGT 180
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
 Db 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80

Db 181 CGAGGCACTCGGGAACCTCGTGGGTGACACTTCTAGAGAGGCTCTCGGTGAGAGACGAG 240
 QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
 Db 241 GACGACTTCCCTTAGAAGTGGCTCTGCGCTGAGAAAGTTCACAGCTTAGAGTAGTGA 300
 QY 101 AspGluLysMetGluArgValSerGluProCysProGlySerLysGluLeuLys 120
 Db 301 GATGAGAAAGAAATGAGAGACTGAGAGAACTGGCCAGGCCCTCCAAAGAGCACTCAAG 360
 QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGluGluTrpThrPhe 140
 Db 361 TTGGAAGAGCTACAGTGTGATGTCTGTGAGAGAGACAGCCGGCAAGGTGACTTTC 420
 QY 141 ThrLeuTrpAspPheAspAsnGlyLysValThrArgGluAspIleThrSerLeuLeu 160
 Db 421 ACTTATATGACTTCGACCAACATGGCAAGTACCCGTGAGACATTTACAGCTTGCTG 480
 QY 161 HisThrIleTrpGluValAlaAspSerValAsnHisSerProThrSerSerLysMet 180
 Db 481 CATACCATCTATGAAGTGTGACTCTCTGTGACCATTCCTCCCATCAAGCAAGACACA 540
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
 Db 541 CTGGGGGTGAAGCTCACCGTGGCTCTGACGGGAGCCAGTAAGAGAGAGCGTCTTTTC 600
 QY 201 AsnGlnAlaAspLeuGlnSerLysArgProArgAlaGluThrLysProThrGluAspLeu 220
 Db 601 AACCATACCATCTGACAGACACAGAGCCCGGAGCAACAAACCGCTGAGAGACTG 660
 QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnLysAspSerArgLeuGlu 240
 Db 661 CTGGCTGGGAGAAAG 720
 QY 241 GlnSerGlyCysTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisTrpLeu 260
 Db 721 CAGCCAGACTGCTACCACTTCGTGGATGAGAAACATTGAGAGAGAAACCACTACTTA 780
 QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
 Db 781 GACCTGGGGAGAGAGAACTACACTCTCAGTTGGACGGGAGATCCCTTGCTGATGCC 840
 QY 281 GlnLysSerGluLeuProProArgTrpSerAspProThrArgSerArgSerHisGluPro 300
 Db 841 CAGAGTCAGAGCTGCCCTCCGATCTCAACCCCACTGCTGCTCTCCAGAGACCA 900
 QY 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
 Db 901 GAAGCTGCCACATCCACACCCGAGGCCCAAGGTGTGAGCCAGGCTCTCTCCACTTC 960
 QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
 Db 961 CTGTGACCCCATTTGGCAAGGCACTCAGACTCCAGCAAGGCTCCGGGCACTCAGAGAT 1020
 QY 341 GlySerLysIlePheValArgSerProLysAlaGlnLysLysSerValGlyValGlyHis 360
 Db 1021 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGAAACATGGATAGGCCAC 1080
 QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 1081 GGGGCCAGAGGTCCAGAAACCAAGCCCTCCACTGTATCCCAACCCATACCTGTCTCCC 1140
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 Db 1141 TGTGCCATGTGGCCACACAGCCAGCCCTTCTCCCACTGTGGACCCCTGGGGCAAG 1200
 QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 Db 1201 AAACACAAGCATCGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 421 AlaSerGlyGly--ProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 439
 Db 1261 GCTGCAAGAGGCTTCACCGTCATGGGGCGGAGACAGGTGAGGAGCTGCGCTGGTGGT 1320

QY 440 ValTYRGIuSerGIuAlaGIuInProValGIuNArgHISGIuHISGIuHISGIu 455
 DB 1321 GTGTACGAGCGACGGCTGGCGAGCCGCTCCAGACGAAACCATCCACACAGGAA 1380
 QY 460 HIGHSISHSISHTYRHSISHSISHTYRGIu 469
 DB 1381 CATCACCAACCATTTATCATCACTTCTATCATG 1410

RESULT 4
 AAA63924 standard; DNA; 1731 BP.
 AAA63924;
 AC 04-DEC-2000 (first entry)
 XX DNA encoding protein related to Drosophila naked cuticle polypeptide.
 DE Nkd gene, naked cuticle gene; segment-polarity gene; wnt signalling;
 KM gene therapy; Nkd defect; cancer; 89.
 XX Mus musculus.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 140..1555
 FT /*tag= a
 FT /product= "naked cuticle polypeptide"
 MO200049034-A1.
 XX 24-AUG-2000.
 XX 17-FEB-2000; 2000MO-US004188.
 XX 17-FEB-1999; 99US-0120646P.
 XX (STRD) UNIV LELAND STANFORD JUNIOR.
 XX Scott M, Zeng W, Wharton K;
 PI
 XX
 DR WPI; 2000-571967/53.
 P-PSDB; AAB08215.
 PT An isolated nucleic acid molecule useful for analyzing (genetic
 PT predileposition to) a disease state and for therapeutic purposes e.g.
 PT treatment of cancer comprises a sequence encoding a naked cuticle
 PT protein.
 XX
 PS Claim 3; Page 43-45; 58pp; English.
 XX
 CC The present sequence encodes a protein related to the Drosophila Nkd
 CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
 CC gene whose expression is induced by wnt signalling. The Nkd polypeptide
 CC acts to antagonize wnt signalling. Nkd may link ion fluxes to the
 CC regulation of wnt signal potency, duration or distribution. The Nkd
 CC polynucleotides can be used for identifying homologous or related
 CC proteins, to modulate the expression or function of Nkd polypeptides, and
 CC in studying associated physiological pathways. Nkd polynucleotides can
 CC also be used in gene therapy to treat disorders associated with Nkd
 CC defects. They may also be used for therapeutic purposes e.g. treatment of
 CC cancer
 CC
 SQ Sequence 1731 BP; 414 A; 521 C; 504 G; 292 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,37e-127 Length: 1731
 Score: 2174.50 Matches: 407
 Percent Similarity: 90.64% Conservative: 19
 Best Local Similarity: 86.60% Mismatches: 43
 Query Match: 87.01% Indels: 1
 DB: 3 Gaps: 1

US-09-993-966-7 (1-470) x AAA63924 (1-1731)
 QY 1 MetGIYLYLeuHISerLYSProAlaAlaValCYLeValaArgaGIuSerProGIuGIY 20
 DB 140 ATGGGGAACCTTCACCTCGAAGCCGGCCGCTGGCAAGGCGAGAGAGAGAGAGAGT 199
 QY 21 AspSerTheAlaValSerAlaAlaATPAlaArguGIY1LeGIuIUTP1LeGIYArg 40
 DB 200 GACAGCTTCTCTGTAAGCGCTGTGGCAAGAAAGCATCAAGAGTGATCGGAGG 259
 QY 41 GlnArgCYSProGIuGIYValISerGIYProArgGIuLeuArgLeuAlaGIYThr1LeGIY 60
 DB 260 CAGCGCTGTCCAGGACAGCGCTTCAGAGACCCCGCTCAGCTGATGGAGGACCTGGT 319
 QY 61 ArgSerThraArgLeuValaGIYAspValaLeuArgAspThraLeuSerGIuGIuGIu 80
 DB 320 CGAGGCACTCGGGAACCTCGTGGGTGACACTTCTAGAGAGAGCTCTCGGTGAGAGAGAG 379
 QY 81 AspAspPheArgLeuGIuValaAlaLeuProProGIuLYSThrAspGIYLeuGIYSerGIY 100
 DB 380 GACGACTTCCCTTAGAAGGTGGCTGCGCTGAGAAAGATCGACACCTAGGTATGTGA 439
 QY 101 AspGIuLYSPMetGIuArgValISerGIuProCYSPProGIYSerLYSPGIuLeuLYS 120
 DB 440 GATGAGAAAGAAATGAGAGACTGAGCGAACCCTGGCCAGGCTTCAGAAAGACCTCAG 499
 QY 121 PheGIuGIuLeuGIuCYAspValaISerMetGIuGIuAspSerArgGIuIUTPThrPhe 140
 DB 500 TTGGAAGACTACAGTGTGATGTCTGTGAGAGAGAGACCGGCAAGGTGACTTTC 559
 QY 141 ThrLeuTYrAspPheAspAlaLeuValaThraArgGIuAsp1LeThrSerLeuLeu 160
 DB 560 ACTCTATATGACTTCCACCAATGGCAAGTACCGTGAGAGACATTACAGCTTGCTG 619
 QY 161 HisThr1LeTYrGIuValaIAspSerSerValaAsnHisSerProThraSerSerLYMet 180
 DB 620 CATACCATCTATGAAGGTGACTCTCTGTGAACATTCCCCACATCAAGCAAGACA 679
 QY 181 LeuArgValaLeuThraValaIAspAspGIYSerGIuSerLYSPAspSerValaLeuVal 200
 DB 680 CTCGGGTGAAGCTCACCGGTGCTCTGACGGAGGACAGGTAAAGAGAGAGCTCTTTC 729
 QY 201 AsnGIuAlaAspLeuInSerAlaArgProArgAlaGIuThraTYrSPProThraGIuAspLeu 220
 DB 740 AACCATACCATCTGAGAGCAAGGCCCGGAGACACCAACCCGCTGAGAGACTG 799
 QY 221 ArgSerTrogIuLYSPGIuArgAlaProLeuArgPheGIuGIYAspSerArgLeuGIu 240
 DB 800 GTGGCTGGAGAAAG 859
 QY 241 GlnSerGIYCYSTYrHISHSISCYValaAspGIuAsn1LeGIuArgaArgaAsn1STYrLeu 260
 DB 860 CAGCCAGACTGCTACACCACTTCGCGATGAGAACTTAGAGGAGAAACCACTACCTA 919
 QY 261 AspLeuAlaGIY1LeGIuLeuTYrThraSerGIuPheGIYProGIYSPProSerValaAla 280
 DB 920 GACCTGGCGGGGTAAAGAACTACAGCTCAAGTTTGACCGGGAACTCCCTTGGTGGCC 979
 QY 281 GlnLYSPSerGIuLeuProProArgThraSerAspProThraArgSerArgSerHISGIuPro 300
 DB 980 CAGAGGTCAAGGTGCCCTCGAATCTCCAAACCCCACTGCTGCTGCCACAGAGCCA 1039
 QY 301 GlnAla1LeHISHSISLeProHISArgLYSPProGIuGIYValaAspProAlaSerPheHISPhe 320
 DB 1040 GAGGCTGCCACATCCACACCGGAGGCCCAAGGTGGAGCCAGGCTCTCCACACTC 1099
 QY 321 LeuAspThraPro1LeAlaLYSPValaSerGIuLeuGIuArgaArgaGIYThraGIuAsp 340
 DB 1100 CTTGACACCCCATTTCTCCAAAGCATCAAGCTCCACCAACGCTCCGGGGCACTCAGAT 1159
 QY 341 GLYSerLYSPHISPheValaArgSerProLYSPAlaGIuGIYLYSPSerValaGIYValaGIYHIS 360
 DB 1160 GGGAGCAAGCACTTTGTGAGGTCCCAAGGCCCAAGGCAAGAACATGGGTATGGGCCAC 1219

QY 361 ValAlaArgGlyAlaAArgAenLyAProLeuGlyProAlaIleProAlaValSerPro 380
 |||||
 Db 1220 GGGGCCAGAGGTGCAAGAGCAAGCTCCACTGGTACCACCACCATCTGTCTCCCCC 1279
 QY 381 SerAlaHisLeuAlaAAserProAlaLeuLeuProSerLeuAlaProLeuGlyHisLyA 400
 |||||
 Db 1280 TCTGCCATCTGGCCACAGCCAGCCCTTCTCCCAACCTCGGACCCCTGGGACACAG 1339
 QY 401 LysHisLySerHisAArgAlaLyGlySerGlnGlnLyCyAArgGlyLeuGlnAlaProLeu 420
 |||||
 Db 1340 AAACAACAGACATTCAGCCAGCAAGAGACGAGCGCTGCCGGCCCTGGCAGGGCCCCCTG 1399
 QY 421 AlaSerGlyLy---ProValLeuGlyAArgGlyHisLeuAArgGlyLeuProAlaLeuVal 439
 |||||
 Db 1400 GCTGCAGAGAGCTCCACCGTCATGAGGCGAGAGCAAGTGAGAGAGCTGCTGCCGTGATG 1459
 QY 440 ValTyGlySerGlnAlaGlyGlnProValGlnAArgHisGlyHisHisHisGly 459
 |||||
 Db 1460 GTGTACGAGAGCCAGGCTAGGCAAGCCGTCAGAGACACGAACCATTCACACAGAA 1519
 QY 460 His 469
 |||||
 Db 1520 CATCACCACTATTCACCACTCTCATCAG 1549

RESULT 5
 AAL51512
 ID AAL51512 standard; DNA, 1401 BP.

AC AAL51512;
 DT 24-APR-2003 (first entry)
 XX

DE Mouse Nkd protein gene sequence.

XX Mouse; gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;
 KM colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
 KM murine.

XX Mus sp.

OS WO200292832-A2.

PN 21-NOV-2002.

PD 27-NOV-2001; 2001WO-US044092.

PR 27-NOV-2000; 2000US-0252884P.

PR 16-MAY-2001; 2001US-0291109P.

PR 01-OCT-2001; 2001US-0325571P.

XX (CHIR) CHIRON CORP.

XX Rohan M, Chan V, Yan D;

XX WPI; 2003-129303/12.

PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
 PT useful for treating cancer involving aberrant Wnt signalling, e.g. colon
 PT cancer, head and neck cancer, ovarian cancer, or breast cancer.

XX Example 2; Fig 4; 99pp; English.

XX The invention comprises the amino acid and coding sequence of the human

CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA

CC and protein sequences are useful for the treatment of cancer involving

CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian

CC cNkd gene and breast cancer). The present DNA sequence represents the mouse

XX Sequence 1401 BP; 336 A; 430 C; 416 G; 219 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,49e-127 Length: 1401
 Score: 2171.50 Matches: 406
 Percent Similarity: 91.01% Conservative: 19
 Best Local Similarity: 86.94% Mismatches: 41
 Query Match: 86.89% Indels: 1
 DB: 10 Gaps: 1
 US-09-993-966-7 (1-470) x AAL51512 (1-1401)

QY 1 MetGlyLyLeuHisSerLySProAlaAlaValCyAlyAArgAArgGlySerProGly 20
 |||||
 Db 1 ATGGGAAACCTTCATCTCGAAGCCGGCCGCTGTGCAAGGCGAGGAGACCCGGAAGT 60
 QY 21 AspSerPheAlaValSerAlaAlaTPAlaAArgLySerGlyLeuGlyLeuTPAlaArg 40
 |||||
 Db 61 GACAGCTTCTCTGAACCTGCTCTGGGCAAGAAAGGCAATCGAGAGTGCATCGGAGG 120
 QY 41 GlnArgCyAProGlyGlyValSerGlyProArgGlnLeuAArgLeuAlaGlyThrIleGly 60
 |||||
 Db 121 CAGGCTGTCCAGGACAGCGTCTCAGAGACCCGTCAGCTGAGATTGGCAGGCACTGGT 180
 QY 61 ArgSerThrAArgGlyLeuValGlyAAspValLeuAArgAAspThrLeuSerGlyGlyGly 80
 |||||
 Db 181 CGAGGCATCTGGGAACTCGTGGGTGACACTTGAAGAGGCTCTGGGTGAGAGAGAG 240
 QY 81 AspAAspPheAArgLeuGlyValAlaLeuProProGlyLeuThrAAspGlyLeuGlySerGly 100
 |||||
 Db 241 GACGACTTCCCTTGAAGATGGCCCTGCTCCCTGAGAAAGATCGACAGCTAGGTAGTGA 300
 QY 101 AspGlyLyLySerMetGlyAArgValSerGlyProCyAProGlySerLySglnLeuLyS 120
 |||||
 Db 301 GATGAGAAAGAAATGAGAGATGAGAGACTGAGCACTGGCCAGGCTCCAAAGAGAGCTCAAG 360
 QY 121 PheGlyGlyLeuGlnCyAAspValSerMetGlyGlyLeuAAspSerArgGlyGlyThrPhe 140
 |||||
 Db 361 TTGGAAGGTACAGTGTGATGCTCTGTGAGAGAGACCGCGCAAGGTGACTTTC 420
 QY 141 ThrLeuTyAAspPheAAspAAsnGlyLyValThrArgGlyLeuAspIleThrSerLeu 160
 |||||
 Db 421 ACTCTATATGACTTCGACCAATGCAATGGCAAGTACCCTGAGACATTCACAGCTTCTG 480
 QY 161 HisThrIleTyGlyValValAAspSerSerValAAsnHisSerProThrSerSerLyMet 180
 |||||
 Db 481 CATACCATCTATGAGAGGTGACTCTCTGTGAACCATTCGCCCATCAAGCAAGACA 540
 QY 181 LeuAArgValLySerThrValAlaProAAspGlySerGlnSerLySAspSerValLeuVal 200
 |||||
 Db 541 CTGGGGGTGAAGCTCACCGTGTCTCTGACGGAGCCAGAGTAAAGAGAGCGTCTTTC 600
 QY 201 AspGlnAlaAAspLeuGlnSerAlaArgProArgAlaGlyThrLySProThrGlyAAsp 220
 |||||
 Db 601 AACCATACCATCTGCGAGACACAGGCCCCGAGACACACCAACCCGCTGAGAGACTG 660
 QY 221 ArgSerTrpGlyLySglnAArgAlaProLeuAArgPheGlnLyAAspSerAArgLeuGly 240
 |||||
 Db 661 CGTGGCTGGAGAAAGAGCAGCGAGCCCACTCAGGTTCCAGAGGTGACAGCCACTGAG 720
 QY 241 GlnSerGlyCySlyThrHisLyCyAValAAspGlyLeuAsnIleGlyAArgAAsnHisLyLeu 260
 |||||
 Db 721 CAGCAGACATGCTACCACTTCGTGGATGAAGAACTTGAAGAGAAACCACTACTTA 780
 QY 261 AspLeuAlaGlyIleGlyAAsnTyThrSerGlnPheGlyProGlySerProSerValAla 280
 |||||
 Db 781 GACCTGGCGGAGATAGAACTACACTCTCACTTTGACCGGATCCCTTCGATGGCC 840
 QY 281 GlnLySerGlyLeuProProArgThrSerAAspProThrAArgSerAArgHisGlyPro 300
 |||||
 Db 841 CAGAAAGTCAAGCTGCCCTCGAATCTCAACCCCACTGCTCTGCTCCCAAGAACCA 900
 QY 301 GlnAlaIleHisIleProHisAArgLySProGlnGlyValAAspProAlaSerPheHisPhe 320
 |||||
 Db 901 GAAGCTGCCACATCCACACAGAGGCCCCAAGGTGAGACCCAGGCTCTTCACCTC 960

QY 321 LeuapThrProIleAlaValSerGluLeuGlnIleArgLeuArgGlyThrGlnAsp 340
DB 961 CTTGACACCCCATTTGCGAAGCATCCAGACGGCTCCGGGCACTCAGAGAT 1020
QY 341 GlySerIleHisPheValArgSerProIleAlaGlnIleYleSerValGlyValGlyHis 360
DB 1021 GGGAGCAGACCTTTGTGAGTGTCCCCCAAGGCCCAAGGCAAGAACATGGGTATGGGCCAC 1080
QY 361 ValIleArgGlyIleArgAnIleYleProIleuGlyProIleAlaIleProIleValSerPro 380
DB 1081 GGGGCGCAGAGGTGCAAGAACAGCTCCACTGGTACCACCACTACTGTCTCTCCCC 1140
QY 381 SerAlaIleuAlaIleSerProAlaLeuIleProSerIleuAlaProIleuGlyHisIleYle 400
DB 1141 TGTGCCATCTGGCCACACAGCCCAAGCCCTTCTCCCACTGGCAGCCCTGGGGCAGCAG 1200
QY 401 LysHisIleYleHisArgAlaYleGluSerGlnGlnIleYleYleArgGlyIleuGlnIleProIleu 420
DB 1201 AAACACAGCATGAGCCAGGACAGAGCCAGCCAGCTCCGGGGCTGCGAGGCCCTCTG 1260
QY 421 AlaSerGlyGly--ProValLeuGlyArgGluHisIleuArgGluIleuProIleVal 439
DB 1261 GCTGACAGAGGCTCCACCGCTCATGGGGCGGAGAGGTGAGGAGCTGCTGCCGTGGTG 1320
QY 440 ValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisIleHisIleGlu 459
DB 1321 GTGTACGAGAGCCAGGCTGGGAGCCGCTCAGAGACACACATCACCACACCA 1380
QY 460 HisHisHisHisIleYleHisHis 466
DB 1381 CATCACCATTTATCACAC 1401

RESULT 6

AAL03214 standard; DNA; 2379 BP.

AAL03214;

21-NOV-2001 (first entry)

Human reproductive system related antigen DNA SEQ ID NO: 5902.

Human; reproductive system related antigen; reproductive system disorder;
cancer; gene therapy; ds.

Homo sapiens.

WQ200155320-A2.

17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.

[illegible]

Alignment Scores:	
Pred. No.:	1.57e-57
Score:	1067.00
Percent Similarity:	58.08%
Best Local Similarity:	51.30%
Query Match:	42.70%
DB:	4
US-09-993-966-7 (1-470)	x AAL03214 (1-2379)

QY	29	TPPLAARGLYSGIYLIEGLUGLUTRPILEGIYARGGLNARGCYSPROGILYGLYVALSER	48
DB	82	TGGCGCATGAGGAGGCGCTGGCTTGACGCGGCGCGCGCTGGCGGTGGTGTTCAC	141
QY	49	GLYPROARGGLNLEUARGLEUALAGLYTHRILIEGLYARGSERTHRARGGLULEUVALIGLY	68
DB	142	TGCTAACCCCGAGCTCGGTGGTAAGACACTATTATGGGACAGGTCAAGAA---CTTCTTGG	198
QY	69	ASP-ValleuARGASPThrIleuSERGIUGLUGLUAASPASPheARGLEUVALA	88
DB	199	GAAGTAGACGTTAAATATGGTTTCAAAATTAAG-----TAATACATGACGAAATCGA	252
QY	88	aleuPROProGLULYsThrASPGLYLeuGLYserGLYASPGLULYsIleuMetGLUARGVA	108
DB	253	ACAAATTCATATAGTATCTAAAGCGTTAAGACCCAAATGACAAACACAGTTTGTGTGCC	312
QY	108	ISerGLUPROCYs-----ProGLY-----	114
DB	313	ATCCACAGCCCAACCAAGCCCTGTTCCTCCGGGGCAACCACTTGACCTCTCGCTGCTTTC	372
QY	115	-SerIleYseGLINleuLYsPheGLUGLULeuGLINCysASPVALserMetGLUleuASPse	134
DB	373	TTCACAGCAAGAAGCTCGGTTCCTTAAGTACAGACACTACGTGTCTTTA-----	421
QY	134	rARGINGLUTRThrPheThrIleuTYrASPpHeASPasnGLYsVALThrARGI	154
DB	421	-----	421
QY	154	uASPILethSERleuMetIleTHRILeTYrGLUVALASPserSerVALasnIle	174
DB	422	-----TGAGCTTGACATATATGTGACTCTTGCTGCATATTTCAC	459
QY	174	rPROThSer-----	177
DB	460	TCCCAACAACCTCCCTTCCCTTCTCTGTCAATAGTTTAATCCTCATAGACTACTCT	519
QY	178	-----SerIleuMetleuARGVALYsleuThrVALAlaPROASP-----	190
DB	520	GTATGTTTATGCAATATACCTTAGAGTT---CTTCTTGCGCCACGATACCTTAGATAGTT	576
QY	191	-----GlySerGLINserTYrASPserVALleuVA	200
DB	577	TTCGGGGCTCTGCTGCCATCTGTCTTCTGCGCTCCTCGGAG--AGATATTTT--	631
QY	200	IASngIuNALASPleuGLINserIleAARGProARGAGIAGIuThrLYsPROThrGLUASPle	220
DB	632	-----GAAGACCTA-----CCAAATAACAGTCTTCCCA-----CT	663
QY	220	uARGserTRPILULYs-----LYsGLNARGAlaPROleuARG-----	232
DB	664	TAGCAGATGGAAAGTTCCATTTACAGAGAGTTTCTGCAAGGACGAGGAGGCGACGTACG	723
QY	223	-----PheGLINLYASPserIARGleuGLINserGLYCysERYrthIshIscYsVALProGL	251
DB	724	GGCATTGGGGGAGTGCCTGTGTTTGGAGAACTGGGGCG-----GGGAGATGTC	774
QY	251	uASNileGLUARGARGasnHISrYrleuASPleuVALAGIYLIEGLUasnTYrTHRserGL	271
DB	775	TGGGGTATAGCGCAAGCCCAAGCAC-----ACAGTAGGTCTCTAATTAATGTCTGTGA	828
QY	271	nPheGLYPro-GLYserPROserVALAAGIuLYsSERGIuLeuPROProARGThrSERa	291
DB	829	ATMGTTCTTAGGGTCCCTTCCTCGGTGGCCGAGAGTCAAGAACTGCCCCCGGACCTCCA	888
QY	291	snPROThrARGserIARGserHISelUProGLUAlaIleHISleIlePROHISARGLYsPROG	311
DB	889	ATCCCACTCGATTCGCTCCCATAGCGCGGAAGCATTCACATCCACACCGAAAGGCC	948
QY	311	INGLYVALASPProLAsERpheHISpheleuASPrrPROIleALYsVALserGLU	331
DB	949	AAGGGTGTGACCCGCGCTCTTCCATCTTGTACACCCCAATGCCAATGCTCCAGAGC	1008
QY	331	eUGINGIuNARGleuARGGLYThrGLINASPGLYserLYsHISpheVALARGserPROLYsA	351

PR	14-AUG-2000	2000US-02252710
PR	14-AUG-2000	2000US-02254476
PR	14-AUG-2000	2000US-02255757
PR	14-AUG-2000	2000US-02257588
PR	14-AUG-2000	2000US-02257599
PR	18-AUG-2000	2000US-02262719
PR	22-AUG-2000	2000US-02266818
PR	22-AUG-2000	2000US-02268688
PR	23-AUG-2000	2000US-02277182
PR	23-AUG-2000	2000US-02277009
PR	30-AUG-2000	2000US-02282827
PR	01-SEP-2000	2000US-02292849
PR	01-SEP-2000	2000US-02292877
PR	01-SEP-2000	2000US-02293433
PR	01-SEP-2000	2000US-02293446
PR	01-SEP-2000	2000US-02293456
PR	05-SEP-2000	2000US-02295099
PR	05-SEP-2000	2000US-02295133
PR	06-SEP-2000	2000US-02304377
PR	06-SEP-2000	2000US-02304488
PR	08-SEP-2000	2000US-02312428
PR	08-SEP-2000	2000US-02312433
PR	08-SEP-2000	2000US-02312444
PR	08-SEP-2000	2000US-02314133
PR	08-SEP-2000	2000US-02314144
PR	08-SEP-2000	2000US-02320868
PR	08-SEP-2000	2000US-02320881
PR	12-SEP-2000	2000US-02319688
PR	14-SEP-2000	2000US-02323388
PR	14-SEP-2000	2000US-02323398
PR	14-SEP-2000	2000US-02323399
PR	14-SEP-2000	2000US-02324001
PR	14-SEP-2000	2000US-02324011
PR	14-SEP-2000	2000US-02330633
PR	14-SEP-2000	2000US-02330643
PR	14-SEP-2000	2000US-02330655
PR	21-SEP-2000	2000US-02344233
PR	21-SEP-2000	2000US-02344274
PR	25-SEP-2000	2000US-02349977
PR	25-SEP-2000	2000US-02349988
PR	26-SEP-2000	2000US-02354844
PR	27-SEP-2000	2000US-02358343
PR	27-SEP-2000	2000US-02358366
PR	29-SEP-2000	2000US-02363377
PR	29-SEP-2000	2000US-02363388
PR	29-SEP-2000	2000US-02363389
PR	29-SEP-2000	2000US-02363399
PR	29-SEP-2000	2000US-02363707
PR	02-OCT-2000	2000US-02370377
PR	02-OCT-2000	2000US-02370388
PR	02-OCT-2000	2000US-02370399
PR	13-OCT-2000	2000US-02370408
PR	13-OCT-2000	2000US-02393953
PR	13-OCT-2000	2000US-02393976
PR	20-OCT-2000	2000US-02412211
PR	20-OCT-2000	2000US-02412212
PR	20-OCT-2000	2000US-02417855
PR	20-OCT-2000	2000US-02417866
PR	20-OCT-2000	2000US-02417877
PR	20-OCT-2000	2000US-02418088
PR	20-OCT-2000	2000US-02418099
PR	20-OCT-2000	2000US-02418266
PR	01-NOV-2000	2000US-02446177
PR	08-NOV-2000	2000US-02464747
PR	08-NOV-2000	2000US-02464756
PR	08-NOV-2000	2000US-02464777
PR	08-NOV-2000	2000US-02464788
PR	08-NOV-2000	2000US-02465343
PR	08-NOV-2000	2000US-02465354
PR	08-NOV-2000	2000US-02465355
PR	08-NOV-2000	2000US-02465366
PR	08-NOV-2000	2000US-02465377

XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #27885.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX MO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Dirmnac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 XX P-PSDB: ABG27894.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 27885; 103bp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridization probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptides and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1,32e-43 Length: 1743
 Score: 842.00 Matches: 196
 Percent Similarity: 43.71% Conservative: 9
 Best Local Similarity: 41.79% Mismatches: 13
 Query Match: 33.69% Indels: 251
 Gaps: 5
 US-09-993-966-7 (1-470) x AAS92081 (1-1743)
 QY 9 AAlaAlaValCybLysAlaGArgGluSerProGlu----- 19
 Db 12 GCCGCCCTGTGCAAGCCGAGGAGAGCCCGAAGGTAGAGGCGCGCGGCGAGACCTC 71
 QY 20 -----GlyAspse 22
 |||||

Db 72 GGGATGAGACGGGGGAGACACCGGGCGCGAGAACGCCGCCCGCGAGGTGACAG 131
 QY 22 rPheAlaValSerAlaAlaTrpAlaArgLygLyLeGluGluTrpIleGlyArgGlnAr 42
 Db 132 CTTGCGCGTAGAGCGCTGCTGGGCTCGAAGAGGACATGAGATGATCGGAGACAGCG 191
 QY 42 GCGProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgse 62
 Db 192 CTGCGCGGCGGTCTCTCGGAGACCCGACAGCTGCGGTGGCGGACCATAGGCCAGAG 251
 QY 62 rThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluAspAs 82
 Db 252 CACCCGG----- 258
 QY 82 pPheArgLeuGluValAlaLeuProProGluLyThrAspGlyLeuGlySerGlyAspG 102
 Db 258 ----- 258
 QY 102 uLysLyMetGluArgValSerGluProCySProGlySerLyAsGlnLeuLySPhneG 122
 Db 258 ----- 258
 QY 122 uGluLeuGlnCyAspValSerMetGluGluAspSerArgGlnGluTrpThrPheThrle 142
 Db 259 ---ATGCTCTGAGCTCCGG----- 276
 QY 142 uTyrAspPheAspAsnAngLyValThrArgGluAspIleThrSerLeuLeuHisTh 162
 Db 276 ----- 276
 QY 162 rIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLyMetLeuAr 182
 Db 276 ----- 276
 QY 182 GValLyLeuThrValAlaProAspGlySerGlnSerLyArgSerValLeuValaEng 202
 Db 277 -----AACTTGACAGCGGCTTCACGACATCTGCGGAACT 314
 QY 202 nAlaAspLeuGlnSerAlaArgProArgAlaGluThrLySProThrGluAspLeuArgse 222
 Db 315 CGGCGAAGCCCAAGGACG----- 333
 QY 222 rTrpGluLyLeuGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlnse 242
 Db 333 ----- 333
 QY 242 rGlyCySThrHisLeuCySValAspGluAsnIleGluArgArgAsnHisTyrLeuAspLe 262
 Db 333 ----- 333
 QY 262 uAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAlaGlnly 282
 Db 334 -----TCCCTTCCTGCGGCCAGAA 353
 QY 282 sSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluProGluAl 302
 Db 354 GTCAGAACTGCCCCCGGACCTCCATATCCACTCGATCTGCTCCATAGCGGAGAC 413
 QY 302 alleHisIleProHisAlaArgLySProGlnGlyValaAspProAlaSerPheHisPheLeuAs 322
 Db 414 CATTCACATCCCAACGAGAGGCCCAAGGCGTGGACCCGAGCTCTTCCATCTTCTTGA 473
 QY 322 pThrProIleAlaLyValSerGluLeuGlnGlnArgLyArgGlyThrGlnAspGlyse 342
 Db 474 CACCCCAATGCGCAAGGTCTCAGAGCTCCAGCAACGCTCCGAGGACCCAGAGACGGAG 533
 QY 342 rLysHisPheValaArgSerProLySAlaGlnGlyLySValaGlyValaIleValaI 362
 Db 534 CAAGCACTTGTGTAGGCCCCCAAGGCCCAAGGCCCAAGAGTGTG----- 576
 QY 362 aArgGlyAlaArgAsnLySProLeuGlyProAlaIleProAlaValSerProSerAl 382
 Db 576 ----- 576

Oy		382	AHISLEUALAASERPROALBLEULEUPROSERIEUALAPROLEUGLYNHSYLSYSTH	402
Db		577	-----GGCCCCCTCGGACACAAGAAGCA	599
Oy		402	SLVSHIARQALATLALYSGLUSERGINGINGLYCYSAATRGYLLEUGNALAPROLEUALASE	422
Db		600	CANAGACCAGGCCAAAGAGAACCCAGCAGGAGGCTTCGGGGGCTTGCAGGCACACTGGCCTC	659
Oy		422	TGYLGYLPVALLEUGLYXARGGLUHISLEULNRGJLUSEUPPROALALEUVAIVALTyrGI	442
Db		660	AGGTGGCCCTGTCCTCGGGGCGGAGAGCACCTCGGGAGAGCTGCCGCTTGTTGCTGATGA	719
Oy		442	USERGINLAAGLYGINPROVALGln	450
Db		720	GAGCCAGGCCGGGCGAGCCGCCACG	744
RESULT 11				
ID	ADCJ1966			
XX	ADCJ1966	standard; cDNA; 1743 BP.		
AC				
XX	ADCJ1966;			
DT		18-DEC-2003 (first entry)		
XX				
DE		Human novel cDNA contig sequence, SEQ ID NO:2048.		
XX				
KW		Human; diagnostic; drug screening; forensics; gene mapping;		
KM		biodiversity assessment; Parkinson's disease; Alzheimer's disease;		
KW		neurodegenerative diseases; anaemia; platelet disorder; wound; burns;		
KM		ulcers; osteoporosis; autoimmune disease; cancer;		
KW		molecular weight marker; food supplement; antiparkinsonian; nootropic;		
KM		neuroprotective; anti-anemic; anticoagulant; thrombolytic; vulnery;		
KW		antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;		
KM		gene therapy; ss.		
OS				
XX		Homo sapiens.		
PN		WO2003029271-A2.		
PD		10-APR-2003.		
XX				
Pf		24-SEP-2002; 2002WO-US030474.		
XX				
PR		24-SEP-2001; 2001US-0324631P.		
XX				
PA		(HYSE-) HYSEQ INC.		
XX				
PI		Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;		
PI		Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;		
PI		Haley-Vicente D, Drmanac RT;		
DR		WPI; 2003-371981/35.		
DR		P-PDSB; ADC32733.		
XX				
PT		New polynucleotide and polypeptide useful for diagnosing, preventing or		
PT		treating conditions such as neurodegenerative diseases, anemias, platelet		
PT		disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or		
PT		cancer.		
PS		Example 2; SEQ ID NO 2048; 1185bp; English.		
XX				
CC		The invention relates to 971 novel human cDNA sequences (ADC29919-		
CC		ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The		
CC		invention also relates to nucleic acid sequences over 9% identical with		
CC		the novel human cDNAs. The invention additionally encompasses expression		
CC		vectors and host cells comprising a nucleic acid of the invention; the		
CC		recombinant production of a polypeptide of the invention; an antibody		
CC		against a polypeptide of the invention; a method of detecting		
CC		polynucleotides or polypeptides of the invention; and methods of		
CC		identifying a compound which binds to a polypeptide of the invention. The		
CC		invention further discloses methods of preventing, treating or		
CC		ameliorating a medical condition; kits comprising polynucleotide probes		

and/or monoclonal antibodies for carrying out the methods of the invention; methods for the identification of compounds that modulate the expression or activity of the polynucleotide and/or polypeptide; and 767 contig sequences corresponding to the cDNA sequences of the invention (ADC31661-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC3394). The nucleic acids and polypeptides of the invention are useful in diagnostic, drug screening, forensic, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a human contig sequence used in an example of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WFO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Alignment Scores:

Alignment Scores:	
Pred. No.:	1,32e-43
Score:	842.00
Percent Similarity:	43.71%
Best Local Similarity:	41.79%
Query Match:	33.69%
DB:	10
Gaps:	5
Length:	174
Matches:	196
Conservative:	9
Mismatches:	13
Indels:	251

US-09-993-966-7 (1-470) x ADC31966 (1-1743)

QY	9	AlaAlaValCysLysArgArgGluSerProGlu-----	19
Db	12	GCCTCCGTGTGCMAACCGCAGGAGAGCCCGAAAGGTAGGGGCGCGCGGGCCGACACTC	71
QY	20	-----	-----GlyAspSe 22
Db	72	GGGGATGGACCGCGGGGACACCGCGGCGCGGCAGAACGCCCCAGCTACACG	131
QY	22	rPheAlaValSerAlaAlaTrpAlaArgLysGlyIleGluGluTrpIleGlyArgGlnAr	42
Db	132	CTTCCCGCGTGAACCGCTGCCTGGGCTCGGAAGGACATCGAGGAGTGTATCGGAGACAGCG	191
QY	42	gCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGlyArgSe	62
Db	192	CTGCCCGGGCGGTGTCTCGGGAGCCCGACAGCTGCGGTTGGCGGGGACCAATAGGCCGAA	251
QY	62	rThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluLysAspAs	82
Db	252	CACCCCGG-----	258
QY	82	pPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGlyAspG	102
Db	258	-----	258
QY	102	uLysLysMetGluArgValSerGluProCysProGlySerLysLysGluLeuLysPheG	122
Db	258	-----	258
QY	122	uGluLeuGlnCysAspAspValSerMetGluGluLysAspSerArgGlnGluTrpThrPheThrLe	142
Db	259	---ATGTCCTGTGAGCTCCG-----	276
QY	142	uTyrAspPheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisThr	162
Db	276	-----	276
QY	162	rIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLysMetLeuAr	182
Db	276	-----	276

```
QY 182 gValValLeuThrValAlaProAspGlySerGlnSerLeuArgSerValLeuValaGlnI 202
Db 277 -----AACTCTGCACGGGGCTTTCAGGACAGCTCGCGGAACT 314
QY 202 nAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeuArgSe 222
Db 315 CGGGCAAGCCCAAGGCGAGC----- 333
QY 222 rTrpGluLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlnGlnSe 242
Db 333 ----- 333
QY 242 rGlyCysTyrHisHisCysValaAspGluAsnIleGluArgArgAsnHisTyrLeuAspLe 262
Db 333 ----- 333
QY 262 uAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValaIaGlnIly 282
Db 334 -----TCCCTTCCTGGGCCAGAA 353
QY 282 sSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluProGluAl 302
Db 354 GTCAGAACTGCCCCCGGACCTCCATCCCATCTCGCTCCATAGCCGGAAGC 413
QY 302 aIleHisIleProHisIleArgLysProGlnGlyValaAspProAlaSerPheHisIleLeuAs 322
Db 414 CATCCCATCTCCACACCGAAAGCCCCCAAGCGTGGACCCGCTTCCTCCACTTCCTTGA 473
QY 322 pThrProIleAlaLysValaSerGluLeuGlnIleArgLeuArgGlyThrGlnAspGlySe 342
Db 474 CACCCCATGTCAGAGCTCTCAAGCTCCAGCAAGGCTCCGGGGACCCAGAGCGGGAG 533
QY 342 rLysHisIlePheValaArgSerProLysAlaGlnGlyLysSerValGlyValaGlyHisValaI 362
Db 534 CAAGCACTTGTGTAGGCGCCCAAGGCCCAAGGCGAAAGTGTG----- 576
QY 362 aArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSerAl 382
Db 576 ----- 576
QY 382 sHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlnIleHisLysValHis 402
Db 577 -----GCCCCCTCGGSCACAAAGCA 599
QY 402 sLysHisIleArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAlaSe 422
Db 600 CAAGCAACGAGCCAGAGAGACGAGCGGCTGCGGGGCTCGAGGACACACTGGCTC 659
QY 422 rGlyGlyProValLeuGlnValArgLysIleLeuArgLysLeuProAlaLeuValValTyrG 442
Db 660 AGGTGGCCCTGTCTCTGGGGGCGAGACCTGCGGGAGCTCGCGCTGTGTGTATGA 719
QY 442 uSerGlnAlaGlyGlnProValGln 450
Db 720 GAGCCAGGCCGGGACAGCGGCCAG 744

RESULT 12
AAS27117
ID AAS27117 standard; cDNA, 1966 BP.
XX AAS27117,
XX
XX 07-NOV-2001 (first entry)
XX
XX cDNA encoding novel signal transduction pathway protein, Seq ID 152.
XX
XX Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;
XX antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
XX immune system disorder; rheumatoid arthritis; inflammatory condition;
XX organ transplant rejection; infection; hepatitis C; blood disorder;
XX sickle cell anemia; hyperproliferative disorder; Gaucher's disease;
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
```

```
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; se;
KW acquired immune deficiency syndrome.
XX
XX Homo sapiens.
XX
XX WO200154733-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001312.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205151P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225477P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 18-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226682P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
```

PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-02344274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM,
XX WPI; 2001-465460/50.
DR P-PSDB; AAU17200.
XX
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders.
XX
PS Claim 1, SEQ ID NO 152; 880pp; English.

XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative disorders
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC (Down syndrome), ischemic injury (e.g. stroke), renal disorders (e.g.
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC respiratory disorders, dermatological disorders, in wound healing,
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders, gastrointestinal disorder
CC (inflammatory disorders), liver disorders, activators of T-cells, to induce
CC B-cell responsiveness to pathogens, activators of T-cells, to induce
CC higher affinity antibodies, and as a means to induce tumour proliferation
CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
CC AAS27850 represent novel signal transduction pathway protein coding
CC sequences and PCR primers of the invention
XX

Alignment Scores:
Pred. No.: 6,91e-43 Length: 1966
Score: 831.50 Matches: 218
Percent Similarity: 51.74% Conservative: 49
Best Local Similarity: 42.25% Mismatches: 142
Query Match: 33.27% Indels: 107
DB: Gaps: 19

US-09-993-966-7 (1-470) x AAS27117 (1-1966)

QY 1 MetGlyLeuHisSerLySProAlaValCyS---LySArgArgGluSerProGlu 19
Db 252 ATGGGGAACCTGCAGTGCAGACGCCGCCGCCGCAAGGGAGGAGAGGCCGGA 311
QY 20 -----GlyAspSerPheAlaValSerAla---Ala 28
Db 312 GGTTCCTCCGGCGCTCCGCCCGGACCGCAGGGGACGCTTCGTGGCTCCGCTACGCC 371
QY 29 TrpAlaArgLySgIleGluGluTrpIleGlyArg-----GlnArgCySPro 44
Db 372 AGCGCGCCGCAAGGCGCGGAGGAGGAGCGGCGCGCGGAGCAAGGAGCTCCCC 431
QY 45 GlyGly-----ValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSer 62
Db 432 AATGGGAGCCCAAGAGAGGGGCTTCCTCCG----- 461
QY 63 ThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluAspAsp 82
Db 462 -----GAGGACCGAG 470
QY 83 PheArgLeuGluValAlaLeuProGluLySerThrAspGly----- 96

Dh 471 TGTCCCTACAGTGGACCTCCCGCTGAGAACTGAGGGCCCGGACACCCGGACAA 530
Qy 97 - - - LeuGlySerGlyAap - - - GlnLysLysMetGlnArgValSerGluProCysArgProGly 114
Dh 531 CTCCTCAGCCAGATGACGAGAGAGGAGGACAAACCGGAGGGCCCGGACGACCGGGC 590
Qy 115 SerLysLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 134
Dh 591 GGGGAGGGC - - - CTCAACTTGAAGCAGCATCCAGTGCCTGAGGAGAGAGAGAGC 647
Qy 135 ArgGlnGlnTyrThrPheThrLeuTyrAaspPheAaspPheAaspGlnLysValThrArgGln 154
Dh 648 CGCCAGAGAGTGAAGCTTACAGCTCTATGACTTTGACAACTCGGAAAGGTACCAAGGAG 707
Qy 155 AspIleThrSerLeuLeuLysIleThrIleTyrGlnValValAaspSerSerValAaspHisSer 174
Dh 708 GACATGTCCAGCCCTCATGACACACATCTATGAGGTCTGATGCTCGGTCAACACTCC 767
Qy 175 ProThrSerLysLysMetLeuArgValLysLeuThrValAlaProAapGlySerGlnSer 194
Dh 768 TCGGGGACAGCAAGACCCCTCGGTGAAGCTAACCGTCAGCCCTGAGCCCTCCAGCAAG 827
Qy 195 LysArgSerValLeuValAaspGlnAlaAaspGlnSerAlaAapProArgAlaGluThr 214
Dh 828 AGGAAGAGGGGTCTCTCTGCTGGCCAGAGCCGGAGGCCACCTTGCAGATGAGAGGT 887
Qy 215 LysArgThrGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 234
Dh 888 GAACGTGACAGAGGAGCCAGGGGTGCTGACAGAGGTTGTCTGACACGTGACAGAGGCC 947
Qy 235 GlnAaspSerArgLeuGlnGlnSerGlyCysTyrHisIleCysValAlaArgLysLysLys 254
Dh 948 AGTACTGACCCCGACCCCTGCTGGAGCCGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTG 1007
Qy 255 ArgArgAaspHisIleTyrLeuAaspLeuAlaGlyIleGlnLysThrThrSerGlnPheGlyPro 274
Dh 1008 CGCAGAAACCACTACTGCTGACCTGCGCGGATTTAGAACTACACTGCTCAAGATTGCGCCCT 1067
Qy 275 GlySerProSerValAlaGlnLysSerGlnLysProArgThrSerAap - - - ProThr 293
Dh 1068 GGGGCCCCCTCTGTCGAAAGCAAGACAGAGAGCCCGGCGCTGCACTCCAGAGCC 1127
Qy 294 ArgSerArgSerHisGlnProGlnAlaIleHisIleProHisIleGlnProGln - - - 311
Dh 1128 CGGTCCCGCTCCGAGGAGCCAGATACATGCTCCATACACACCCGACAGTCCAGTCTCTG 1187
Qy 312 - - - GlnValAapProAlaSerPheHisPheLeuAapThrProIleAlaLysValSer 329
Dh 1188 GTGGAAACGTCGTGCCAGCTCG - - - - - GACCTGCTGCTCGCGGCGCTG 1232
Qy 330 GlnLysGlnGlnArgLysArgGlyThrGlnAapGlySerLysHisPheValArgSerPro 349
Dh 1233 GACACGAGCCCGCGCGGAGGGGCGGAG - - - - - AAGACGTTCTTCAGTCCGCC 1283
Qy 350 LysAlaGlnLysLysSerValGlnValGlnHisValAlaArgGlnAlaArgLysLysPro 369
Dh 1284 AAGGGCTCCGGAAGCGG - - - - - 1301
Qy 370 ProLeuGlnProAlaIleProAlaValSerProSer - - - - - AlaHisLeuAlaAlaSer 387
Dh 1302 - - - - - CTTGGGGTGCAGCCAGCAGCAAGTCCGGAAAGCTTCACTACTACTG 1352
Qy 388 ProAlaLeuLeuProSerLeuAla - - - - - 395
Dh 1353 CCGGCGGTCTCTCGCGCCAGGCGCTCAGAGACGAGCCACACACCTCCGAGCCCGGACCG 1412
Qy 396 - - - - - ProLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 414
Dh 1413 CCACTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1463
Qy 415 GlnLysGlnAlaProLeuAlaSerGlnGlyProValLeuGlnArgGlnHis - - - - - Leu 432
Dh 1464 CCACTCAAGGCCCAACGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1514

Qy 433 ArgGlnLeuProAlaLeuValValTyrGlnSerGlnAlaGlnGlnProValGlnArgHis 452
Dh 1515 CGGAGCTTGGCGCCCGCCAGCCAGAGAGGAGGTACCGGGTGCAGTATCCAGCGGCAC 1574
Qy 453 GlnHisIleHisIleHisIleHisIleHisIleHisIleHisIleHisIleHisIleHis 468
Dh 1575 GAGCACCACCAACCAAGAGCACCACCAACCAACCAACCAACCAACCAACCAACCAAC 1622
RESULT 13
ADB93295
ID ADB93295 standard; cDNA; 1966 bp.
AC ADB93295;
XX 04-DEC-2003 (first entry)
DE Human cDNA encoding a novel protein #142.
XX ser; gene; human; autoimmune disease; Parkinson's disease; silicosis;
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
KW immunosuppressive agent; adjuvant; enhance immune response;
KW higher affinity antibody induction;
KW increased serum immunoglobulin concentration.
OS Homo sapiens.
XX US2002168711-A1.
XX 14-NOV-2002.
XX 17-JAN-2001; 2001US-00764868.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 26-JUN-2000; 2000US-0214986P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218280P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226688P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.

ADQ22400
ID ADQ22400 standard; DNA; 1959 BP.
XX
AC ADQ22400;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5220.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM ds.
XX Homo sapiens.
XX OS
XX PN NO2004048938-A2.
XX PD 10-JUN-2004.
XX PF 26-NOV-2003; 2003WO-US038193.
XX PR 26-NOV-2002; 2002US-0429739P.
XX PA (PROT-) PROTEIN DESIGN LABS INC.
XX PI Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX Example 2; SEQ ID NO 5220; 210bp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostratic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 1959 BP; 408 A; 677 C; 612 G; 244 T; 0 U; 18 Other;

Alignment Scores:
Pred. No.: 4,84e-40 Length: 1959
Score: 786.00 Matches: 215
Percent Similarity: 47.24% Conservative: 50
Best Local Similarity: 38.32% Mismatches: 152
Query Match: 31.45% Indels: 145
DB: 12 Gaps: 20

US-09-993-966-7 (1-470) x ADQ22400 (1-1959)
QY 1 MetGlyLysLeuHisSerIysProAlaAlaValCys--LysArgArgGluSerProGlu 19
DB 98 ATGGGAAATGCACTGCAGACACCGCCGCCGCCGCAAGCGAAGAGAGCCCGAA 157
QY 20 GlyAspSerPheAlaValSerAla---AlaTrpAlaArgLysGlyIleGluGluTrpIle 38
DB 158 GGGGACAGCTTCGTCGCTCCGCGTACCGAGCGCGCCGCAAGCGCGAGAGCGGAG 217
QY 39 GlyArg-----GlnArgCysProGlyIly-----ValSerGlyProArgGln 52
DB 218 CGGCGCGCGCGGAGCAAGCAGAGAGCTGCCAATGGGAGCCCAAGAGAGGCGCTTTCCGG 277

QY 53 LeuArgLeuAlaGlyThrIleGlyArgSerThrArgGluLeuValGlyAspValLeuArg 72
DB 277 ----- 277
QY 73 AspThrLeuSerGluGluGluAspPheArgLeuGluValAlaLeuProProGlu 92
DB 278 -----GAGGACCAAGTGTCCCTACAGGTGAGCTCCCGCTGAG 316
QY 93 LysThrAspGly-----LeuGlySerGlyAsp---GluLysLys 104
DB 317 AAAGCTAGGGCGCGAGACCCGGAGCAACTCTTACAGCCCAATGACGAGAGAGGCA 376
QY 105 MetGluArgValSerGluProCysProGlySerLysLeuGluLeuLysPheGluGluLeu 124
DB 377 GCAACCGCAGAGGCGCGGAGGAGACCGGGCGGAGAGCGC---CTCAACATTAGCGCACTC 433
QY 125 GlnCysAspValSerMetGluGluAspSerArgGluTrpThrPheThrLeuTrpAsp 144
DB 434 CAGTGCATGTCTCGGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
QY 145 PheAspAspAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisThrIleTyr 164
DB 494 TTGACCAACTGCGGAAAGTCAACAGAGAGACATGTCCAGCTCATGACACCATATTAT 553
QY 165 GluValValAspSerSerValAsnHisSerProThrSerSerLysMetLeuArgValLys 184
DB 554 GAGTGTGATGCTCGTCAACCACTCTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 613
QY 185 LeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuValAsnGlnAlaAsp 204
DB 614 CTAACGCTCAGCCCTGAGCCCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 673
QY 205 LeuGlnSerLysArgProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGlu 224
DB 674 CGGAGCCCAACCGGTGACAGATGAGGAGGTGATGACAGAGAGAGAGAGAGAGAGAG 733
QY 225 LysLeuGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGluGlnSerGlyCys 244
DB 734 AGAAGTGTGTCTGCACACGTCAAGAGGCCAGTACAGCCCAAGCCCTGTGGAGCGG 793
QY 245 TyrHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeuAspLeuAlaGly 264
DB 794 GGGCCCTACTGCTGAGCGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 853
QY 265 IleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAlaGlnLysSerGlu 284
DB 854 ATTGAGAACTTACGTTCAGATTCTGGCTCTGAGTCTGAGTGCAGAAAGAGAGCTCC 913
QY 285 LeuPro----- 286
DB 914 GCTCCAGAGACACAGACTGGGAGCAAGAGCTAGAGAGTCCGCTTGCAGAGAGCTGTGG 973
QY 286 ----- 286
DB 974 AGCAGGAGAGTACACCAAGTGGCCAGGCCCTTCCCTTCAAGGAGTGTGGCCGTCTGA 1033
QY 286 ----- 286
DB 1034 CTGCAGACTTGCTTACAGACTGAGCTCAGAGTCCCTCTGTGCAAGCAAGCAGAGAGC 1093
QY 287 -----ProArgThrSerAsnProThrArgSerArgSerHisGluProGluVal 302
DB 1094 CCAAGGAGAGGCTCTGCACCTTCCAGGCCCGG---TCCGCTCTCCAGAGCCAGATACA 1149
QY 303 IleHisIleProHisArgLysProGln-----GlyValAspProAlaSerPhe 318
DB 1150 CATGCGGTACACACCGCAGAGTCAAGAGTCTGTGTGACACAGTCGTCAGAGCTTCG--- 1206
QY 319 HisPheLeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThr 338
DB 1207 -----GAGCTGTCTCCCGGCGCTGAGACGAGAGAGAGAGAGAGAGAGAGAG 1254
QY 339 GlnAspGlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyVal 358

Dd 1255 GAG-----AAAGCAATTCTCAAGTACCCCAAAGGCGTCGGAGAACGCCCTGGGGNNN 1305
Qy 359 GlyHsValAlaArgGlyAlaArgAsnLysProProlLeuGlyProAlaIleProlAval 378
Dd 1306 NNNNNNNNNNNNNNTCCGGAAA-----GCCTTCAGACTACTACTCCTCGGCGGTTC 1356
Qy 379 SerPro-----SerAlaHisLeuAlaAsrProlAileuLeuPro 392
Dd 1357 CTGGCGGCCCCAGGCGCTCAGAGCGGCACCACCACTCCGCGAGCCGCCACGS----- 1407
Qy 393 SerLeuAlaProlLeuGlyHisIleValbYhisIshArgAlaLysGluSerGlnGly 412
Dd 1408 -----CCACCTTAGCGGCACAGAAGGGATACCGCCAAAGGGCAGAGGA-----GGC 1452
Qy 413 CysArgGlyLeuGlnAlaProlLeuAlaSerGlyGlyProValLeuGlyArgGluHis- 431
Dd 1453 CACTTCGCCACTCMAAGGCCCAACGCTACGCTGCACACATG-----GAGCACAG 1503
Qy 432 --LeuArgGluLeuProAlaLeuValTyrguSerGlnAlaGluProValGln 450
Dd 1504 GTGGTGCGGAGACTGCGCGCCACCGCACAGAGAAGGGCTACGGGGTCCAGTATCCAG 1563
Qy 451 ArgHisGluHisHisHisHis-----HisGluHisHisHisHisHisHisHisHe 467
Dd 1564 CGGCGACGACCAACAACCAACCAACGACGACCAACCAACCAACCAACCAACCACTTC 1623
Qy 468 Tyr 468
Dd 1624 CAC 1626

Search completed: December 30, 2004, 04:02:31
Job time : 874 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM proteoin - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 00:53:19 ; Search time 8160 Seconds
(without alignments)
2723.794 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 2499
Sequence: 1 MGKLSKPAVACGRSPG.....RHHHHHHHHHHHHFYOT 470

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues.
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool.p/US09993966/runat.29122004.130758.8996/app.query.fasta_1.647
-DB=GenEmbl -QFMT=fasta -SUFFIX=xge -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0
-INITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09993966.@CEN.1.1.7406@runat.29122004.130758.8996 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_ba:.*
2: gb_hc:.*
3: gb_in:.*
4: gb_cm:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_st:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vl:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2499	100.0	1438	9 AF358135 Homo sapi
2	2499	100.0	1786	9 AY061883 Homo sapi
3	2499	100.0	1788	9 AB062886 Homo sapi
4	2499	100.0	2604	9 BC051288 Homo sapi

Result No.	Score	Query Match Length	DB ID	Description
5	2493	99.8	1438	6 AR405917 Sequence
6	2192.5	87.7	1416	10 AF343352
7	2191.5	87.7	1765	10 BC034838
8	2189.5	87.6	1671	10 AF358134
9	2174.5	87.0	1401	6 AR405916
10	2167.5	86.7	1401	6 AX172351
11	1677	67.1	1053	6 C0721681
12	1210.5	48.4	207486	2 AC145175
13	1191.5	47.7	163319	9 HSA303140
14	1191.5	47.7	169773	9 AC007608
15	1191.5	47.7	170791	9 AC007334
16	1191.5	47.7	178790	2 AC145238
17	1135	45.4	194869	2 AC145003
18	1135	45.4	198189	2 AC145043
19	1112	44.5	232509	2 AC134093
20	1076.5	43.1	208061	2 AC145254
21	1075	43.0	240965	2 AC126054
22	1055.5	42.2	199277	2 AC145445
23	1029	41.2	139145	2 AC146326
24	939	37.6	181401	2 AC146663
25	872.5	34.9	1967	10 BC019952
26	864.5	34.6	1947	10 AF358136
27	846.5	33.9	1682	9 AF358137
28	846.5	33.9	1798	9 AB062887
29	793	31.7	1958	9 BC012176
30	780.5	31.3	1285	6 AR405918
31	741.5	31.2	1885	6 AX780854
32	641.5	25.7	1309	9 BC004940
33	637.5	25.5	212531	2 AC145178
34	628	25.1	1191	6 C0715890
35	583.5	23.3	1307	6 AR405919
36	314.5	12.6	181105	5 BX664719
37	285.5	11.4	99395	2 AC010446
38	285.5	11.4	187289	9 AC116351
39	272	10.9	162641	2 AC091406
40	272	10.9	208878	2 AC141571
41	272	10.9	246241	2 AC135365
42	246	9.8	1950	6 AX557538
43	240.5	9.6	2340	6 C0741468
44	224	9.0	237332	2 AC131846
45	215.5	8.6	2691	6 C0605931

ALIGNMENTS

RESULT 1
LOCUS AF358135 1438 bp mRNA linear PRI 29-MAY-2001
DEFINITION Homo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.
ACCESSION AF358135
VERSION AF358135.1 GI:14211713
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 1438)
TITLE Wharton, K.A., Jr., Zimmermann, G., Rousset, R. and Scott, M.P.
Dishvelled and antagonize Wnt signaling
JOURNAL Dev. Biol. 234 (1), 93-106 (2001)
MEDLINE 2125530
PUBMED 11356022
REFERENCE 2 (bases 1 to 1438)
AUTHORS Wharton, K.A., Jr., Zimmermann, G. and Scott, M.P.
TITLE Direct Substitution
JOURNAL Submitted (08-MAR-2001) Pathology and Molecular Biology, University
of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,
Dallas, TX 75390-9072, USA
FEATURES
source
1..1438
/organism="Homo sapiens"
/mol_type="mRNA"

gene
/db xref="taxon:9606"
/chromosome="16"
/map="16q12"
1..1438
/gene="NKD1"
8..1420
/note="containing EFX domain; similar to Drosophila melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"

CDS

/codon_start=1
/product="naked cuticle-1"
/protein_id="AAK57484.1"
/db_xref="GI:14211714"
/translation="MGKLSKPAVCKRRSPGSDSPASAAARKGIEEMIGRORCP
GVYSRROLACTIGSRPRLVDYRDLSPRRDPRLFYALPEKRDGSGSDP
KRRERVSERPGSKQKLFEBLOCDVSMEDSROERTFTLYDPNNKRTREDITSL
HTLYEVDSSVNSPTSSKMLRYKLTVAPGSQSKRVLVQADLQASRAPEKPE
DSWEKQKRAPLRFQDSRLQSGCYHHCVDENIERNNHYLDLAGIENTYSGPDS
PSVAQSELPRTSNPRTSRSHPEAIIHRRKQGVDPASFPFLDPIAKVSELQOR
LRGTQSGKHFRVSPKQSGKGVGVHARANKRPGLPAIPVPSAHLAASPLALP
SLAPLGKHKHKHAKESQSGRGIGAPLASGPFVLRHRLRELPAIVYTESQAGPQ
RHEHHHHHHHHHHHYT"

ORIGIN

Alignment Scores:

Pred. No.: 3,17e-112 Length: 1438
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-993-966-7 (1-470) x AFS58135 (1-1438)

QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLysArgArgLysSerProGluGly 20
Db 8 ATGGGGAAACTTCACTCCAAAGCCGGCCGGTGTGCAAGCCAGAGAGAGAGCCGGAAGT 67
QY 21 ArgSerPheAlaValSerAlaAlaTTPAlaArgLysGlyLysGluGluTTPILegLYArg 40
Db 68 GAACAGCTCCCGCGAGCCGCTGCGAGGCGATCGAAGGCGATCGAGAGTGAATGGAGA 127
QY 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLysGluAlaGlyThrIleGly 60
Db 128 CAGCGCTGCCCGGCGGTGTCTCGGAGCCCGCAAGCTGCGGTTGGCGGACCAATAGGC 187
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCCGGGAGCTCGTGGCGGACGTGTGAGAGCACGCTCAAGAGAGAGAG 247
QY 81 AspAspPheArgGluGluValAlaLeuProProGluLysThrArgGlyLeuGlySerGly 100
Db 248 GAGCACTTTGGCGTGAAGTGGCCCTGCTCTGAGAGAGATGACGGGCTGGGAGCGGA 307
QY 101 AspGluLysLysMetGluArgValSerGluProCysProGlySerLysLysGlnLeuLys 120
Db 308 GATGAGAAAGATGAGAGAGTGAAGCAACCTTGGCCAGGCTCAAGAGACACTGAAG 367
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
Db 368 TTGTAAAGACTCCAGTCCGAGTGTCCATGAGAGAGACCGCGGAGAGTGAACCTTC 427
QY 141 ThrLeuTyrrAspPheAspAsnArgGlyLysValThrArgGluAspLysIleThrSerLeuLeu 160
Db 428 ACCCTGATATGACTTTGACAAACGCAAGGTCAACCGAGAGGACATCACCAAGCTTCGT 487
QY 161 HisThrIleTyrrGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 488 CACACCACTATGAGGAGTGAAGTCTCTGTCAACCACTCCCAACATTCAGCAAGATG 547
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200
Db 548 CTGGGGTAAAGCTCACCGTGGCCCGGATGGCAGCCAGAGCAAGAGAGAGTCTTGTCTC 607

QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
Db 608 AATCAGGCTAAGCTTGCAGAGCCGAGAGCCCGGAGAGAGACCAAGCCCATGAGAGACTG 667
QY 221 ArgSerTyrGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGGAGCTGGAGAAAGAGAGAGCCAGCCCGCTCAGGTTCCAGAGGTGACAGCCGCTGGAG 727
QY 241 GlnSerGlyCysTyrrHisSerGlyValAspGluAsnIleGluLysArgAsnHisIleTyrLeu 260
Db 728 CAGCTGGCTGATACCAACCATTCGTATGATGAAGAACTCAGAGAGAGAAACCACTACTTA 787
QY 261 AspLeuAlaGlyLysLeuAsnTyrrThrSerGlnPheGlyProGlySerProSerValAla 280
Db 788 GATCTCCCGGAGATGAAGAACTACACTCCCATTTGGGCTGGCTCCCTCCGTGGCC 847
QY 281 GlnLysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 848 CAGAGTCAGAACTGCCCCCGACCTCCATCCCATCTCGATCTCGCTCCCATGAGCCG 907
QY 301 GlnAlaIleHisIleProHisArgGlySerProGlnGlyValAspProAlaSerPheHisPhe 320
Db 908 GAAGCCATCCACATCCACACCAAGCCCAAGGCGTGAACCGGCTCTCCACTTC 967
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
Db 968 CTTGACACCCCAATCGCCAGAGTCTTCAAGCTTCAGACACGCTCCGGGACACCAAGAC 1027
QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db 1028 GGGAGCAAGCACTTTGTGAGTCCCCCAAGGCCAGGCAAGATGTGGGTGGGCAC 1087
QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1088 GTGGCCAGAGGGCCAAAGAAACAAGCCCTCTGGAGACCCCAATCCCTGGGTGCTCCC 1147
QY 381 SerAlaHisLeuAlaIleSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1148 TCGCCCACTGGGTGCAAGCCCGGCTCTCTCCCTCCCTAGCCCCCTCGGACACAG 1207
QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1208 AAGCACAAGACCGAGCCAGAGCAGAGCAGACAGGCGTCCGGGCGCTGACAGCACCACTG 1267
QY 421 AlaSerGlyGlyProValLeuGlyArgGlnHisLeuArgGluLeuProAlaLeuValVal 440
Db 1268 GCTTCAGGTGGCTGTCTGTGGGGCGGAGACACTGCGGAGGCTGCCGCTTGTTGGTG 1327
QY 441 TyrGlnSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisGluHis 460
Db 1328 TAGAGAGCCAGGGCGGAGCGGCTCCAGAGACATAGACACCAACACATGAACAT 1387
QY 461 HisHisSerTyrrHisHisPheTyrrGlnThr 470
Db 1388 CACCACTTACCAACCACTTCTACAGACA 1417
RESULT 2
AY061883 1786 bp mRNA linear PRI 26-DEC-2001
LOCUS Homo sapiens naked protein (NKD) mRNA, complete cds.
DEFINITION
AY061883
VERSION
AY061883.1 GI:17978536
KEYWORDS
Homo sapiens (human)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1786)
Yan, D., Wisemann, M., Rohan, M., Chan, V., Jefferson, A. B., Guo, L.,
Sakamoto, D., Coothern, R. H., Fuller, J. H., Reinhard, C., Garcia, P. D.,
Randazzo, F. M., Bacchedo, J., Pant, W. J. and Williams, L. T.
Elevated expression of axin2 and nkhd mRNA provides evidence that

Mut/beta -catenin signaling is activated in human colon tumors
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14973-14978 (2001)

JOURNAL
PUBMED
11752446
2 (bases 1 to 1786)
AUTHORS
Rohan, M., Chan, V., Yan, D., Fan, J., W. J. and Williams, L. T.
TITLE
Direct Submission
Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560
JOURNAL
Horton Street, Emeryville, CA 94608, USA
FEATURES
Location/Qualifiers
1. 1786
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. 1786
/gene="NKD"
159. 1571
/gene="NKD"
/note="NKD"
/product="naked protein"
/protein_id="AAJ32374.1"
/db_xref="GI:17978537"
/translation="MGKLHSPKPAVCKRSPGSDPFAVAAWARKGIEEMIGRORCP
KGVSGRDLRAGTIGRSTRELYDYLRLTSEEDPFLVALPEKTDGSGDPE
KMEVSEPCRGSKKOLKREBLQCDVSMEDSROBWTFTYDNNKGTREDITSL
HTIYEVDSVNSPTSSKRLVLTVAFDGSKSVLVNADLVGSARPRATKPTPE
DLKSWKQKAPLRFGDSRLQSGCYHNCVDENIERNNYLDLAGIENTSGFSPS
PSVAQSELPRTSNPTSRSHSPKALHPIHAKPQGVDPSPALPIKVELQOR
LRGTGDSKHFVSPKAGSKSVGVHVARANKPVGPAIPVSPAHIAAPLALP
SLAPLGKHKHAKESQSGCRGLQAPLASGPRVLGRHRELPAVVYSSQAGPRVQ
RHHHHHHHHHHHHYHPTQ"

ORIGIN

Alignment Scores:

Pred. No.: 4.08e-112 Length: 1786
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conserved: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-993-966-7 (1-470) x AY061883 (1-1786)

QY 1 MetGlyValLeuHisSerIleProAlaAlaValCylValArgArgIleSerProGluGly 20
DB 159 ATGGGAACTTCACTCAAGCCGCGCGCTGTGCAAGCGCAGGAGAGCCCGAAGGT 218
QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgIleGluGlyTTPPIleGlyArg 40
DB 219 GACAGCTTGGCGCTGAGCGCTGCGCTGCGAAGGCGCATCGAGAGTGATGATGAGA 278
QY 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
DB 279 CAGCGCTGCGCGCGCGCTGTCTCGGAGCCCGACAGCTGTGCGGCGCACATGCGC 338
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
DB 339 CGAAGACCCCGGAGCTCGTGGCGCAGCTGTGAGAGACACGCTACGCGAGGAAGAG 398
QY 81 AspAspPheArgLeuGluValAlaLeuProGluGluLeuThrArgGlyLeuGlySerGly 100
DB 399 GACGACTTTCGCTGGAAGTGGCCCTGCTCTGAGAGACTAGCGGCTGGCGAGCGGA 458
QY 101 AspGluLeuLeuMetGluArgValSerGluProCysArgProGlySerIleValGlnLeuVal 120
DB 459 GATGAGAGAGATGAGAGAGTGAAGGAGACCTGCGCCAGCTCCAGAGAGAGCTGAG 518
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
DB 519 TTGGAAGAGCTTCAGTGGCGAGCTGTCTATGAGAGAGACCGCGCAGAGTGAACCTTC 578
QY 141 ThrLeuTyrAspPheAspPheAsnGlyValValThrArgGluLeuAspIleThrSerLeuLeu 160

DB 579 ACCCTGATGACTTTGACAAACGCGCAAGTACCCGAGAGGACATCACCAGTTGCTG 638
QY 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerIleMet 180
DB 639 CACACCATCATGAGGGTGGTGAAGCTCTCTGTCAACCATCCCAATCCAGCAAGATG 698
QY 181 LeuArgValIleLeuThrValAlaProAspGlySerGlnSerIleValArgSerValLeuVal 200
DB 699 CTGGGGGTAAAGCTCACCGTGGCCCGCATGGCAGCCAGCAAGAGAGAGGCTCTGTGC 758
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleValProThrGluAspLeu 220
DB 759 AATCAGGCTAAGCTGCGAGAGCGAGGCCCGGAGCAAGACCAAGCCATGAGAGACTG 818
QY 221 ArgSerTyrGluValIleGluArgValAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
DB 819 CGGAGCTGGAG 878
QY 241 GlnSerGlyCysTyrThrHisCysValAspGluAsnIleGluValArgAsnHisTyrLeu 260
DB 879 CAGTCTGCTGTACCAACCATTCGTATGATGAGAACATCGAGAGAGAGAGAGAGAG 938
QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
DB 939 GATCTCGCGGAGTAAAGAACTACACGTCCCAATTGGGCTGGCTCCCTTCGTTGCC 998
QY 281 GlnIleSerGluLeuProProArgTyrSerAspProThrArgSerArgSerHisGluPro 300
DB 999 CAGAGTCAGAACTGCCCCCCCCGACCTCAATCCCATCGATGCTGCCATGAGACCG 1058
QY 301 GlnAlaIleHisIleProHisArgIleProGlnGlyValIleProAlaSerPheHisPhe 320
DB 1059 GAAAGCATTCACATCCACACCAAGGCCCAAGGCGCTGAGACCGGCTCTCCACTTC 1118
QY 321 LeuAspThrProIleAlaValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
DB 1119 CTTGACACCCCAATCCGCAAGTCTCAGACTTCCAGACAGGCTCCGAGGACCCAGAG 1178
QY 341 GlySerIleHisIlePheValArgSerProValGlnGlyIleSerValGlyValGlyHis 360
DB 1179 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGAGTGGGTGGGCGAC 1228
QY 361 ValAlaArgGlyValArgAsnIleProProLeuGlyProAlaIleProAlaValSerPro 380
DB 1239 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGAGACCCGCCATCCCTGGGTGCTCCCC 1298
QY 381 SerAlaHisLeuAlaIleSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisPhe 400
DB 1299 TCCGCCCACTGGCTGCGACGCCGCGCTCTCTCCCTTACGCCCTTCGGGACAG 1358
QY 401 LeuHisIleHisIleArgAlaValSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
DB 1359 AAGCACAAGACCGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1418
QY 421 AlaSerGlyGlyProValLeuGlyValArgGluHisLeuArgGluLeuProAlaLeuVal 440
DB 1419 GCCTCAGGTGGCCCTGTCTGCGGCGGAGACCTGCGGAGAGCTGCGCGCTGTGGTG 1478
QY 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisIleHisIleGluHis 460
DB 1479 TATGAGAGCCAGGCGCGGAGCGGCTCCAGAGACATGAGCACCAACCAATGAACAT 1538
QY 461 HisHisIleTyrHisIlePheTyrGlnThr 470
DB 1539 CACCACATTACCAACCACTTCTACAGACA 1568
RESULT 3
AB062886 1788 bp mRNA linear PRI 20-OCT-2001
LOCUS Homo sapiens mRNA for Dvl-binding protein NKD1, complete cds.
ACCESSION AB062886
VERSION AB062886.1 GI:16303259
KEYWORDS

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1
TITLE Molecular cloning, gene structure, and expression analyses of NKD1 and NKD2
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PUBMED 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh, M.
TITLE Direct Submission
SUBMITTED (08-JUN-2001) Masaru Katoh, National Cancer Center Research Institute, Genetics and Cell Biology Section, Tsukiji 5-chome, Chuo-Ku, Tokyo 104-0045, Japan (E-mail:mkatoh@ncc.go.jp, Tel:81-3-3542-2511)
FEATURES
source location/Qualifiers
1. 1788 /organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="human fetal tissues"
1. 1788 /gene="NKD1"
6. 1418 /gene="NKD1"
/codon_start=1
/product="Dvl-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16103260"
/translation="MGKLSKPAVAVCKRRESPEEDSPVSAAMARKIEEMIGRORCPGVSPRQLAGTIGRSTELVGDVLRDTLSEEDPFLLEVALPEKTDGASGDEKNRVERPCPGSKQLEKFELOCDVMEBDSRQEMFTLYDPNNKAVREDITSLHTLYEVVDSVNHSPITSKMLRYLTVAPGSGSKRSVLVQADLQARPAETKPTSLDSMEKKORAPLRPOGDSRLGOSGCHNCVDENIERNNHYLDLAGIENTSGRPGSPSVYAKSELPEPTSNPFRSRSHPEAHTIHARKQGVDPASPHRLDPTIAKVELTORLRKTODGSKRPVSPRAQGSVGVGHVARKARKNPRIQPIPAVSPAHLLAASALILP SLAPLGHKKHKKHAKESQDGCRLQAPLASGCPVLGKRLRELPAVVESQAGQPVQ RHHHHHHHHHHHHYHNYQT"
ORIGIN
Alignment Scores:
Pred. No.: 4.09e-112 Length: 1788
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservat: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-993-966-7 (1-470) x AB062886 (1-1788)
QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLysArgArgLysSerProGluGly 20
DB 6 ATGGGAAACTTCTCTCAAGCCGCGCCGCTGTGCAAGCCGAGGAGCCCGGAAGT 65
QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyTlleglGluTTPllleglYArg 40
DB 66 GACAGCTTCGCGCTGACGCGCTGCTCGAAGGCGCATCGAGAGTGAGTGGAGAGA 125
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyTThlleglY 60
DB 126 CAGGCGTCCCGGGCGGCTGTCTGGAGACCCGACAGCTGCGGTTGGCGGGGACCAATAGG 185
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
DB 186 CGAAGCACCCGGAGAGCTGTGTGGCGACCTGTGAGAGACAGCTCAAGAGGAAGAGAG 245
QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
DB 246 GACGACTTCGGCTGGAAGTGGCCCTGCTCTGTAAGAAAGTACGAGGCTGGGACGCGA 305

QY 101 AspGluLysMetGluArgValSerGluProCysProGlySerLysGluLeuLys 120
DB 306 GATGAGAGAGAGTGAAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
DB 366 TTGAAGAGCTCCAGTGGCAGCTGTCCATGAGAGAGAGAGAGAGAGAGAGAGAG 425
QY 141 ThrLeuThrAspPheAspAspAspGlyLysValThrArgLysAspLlthrsSerLeu 160
DB 426 ACCCTGATACCTTGAACAACAGGAGAGTACCCGAGAGAGACATCACAGCTTGTCTG 485
QY 161 HisThrLlthrsGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
DB 486 CACACCATCTATAGGTGGTGAAGCTCTCTGTACACACCTCCCAACATCAGAGATG 545
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
DB 546 CTGCGGGTAAAGCTCACCGTGGCCCGGATGGACGACGAGCAAGAGAGAGAGAGAG 605
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
DB 606 AATCAGGCTGACCTGACAGAGCGAGAGCCCGAGCAGAGACCAAGCCACTGAGAGACTTG 665
QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
DB 666 CGGAGCTGGAG 725
QY 241 GlnSerGlyCysTrpThrHisLysCysValAspGluAsnLlleglUArgArgAsnHisTrpLeu 260
DB 726 CAGTGGCTGCTGATCCACCATTCGTGATGAGAAACATCAAGAGAGAGAGAGAGAG 785
QY 261 AspLeuAlaGlyLlleglUAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
DB 786 GATCTGCGCGAG 845
QY 281 GlnLysSerGluLeuProProArgTrpSerAspProThrArgSerArgSerHisGluPro 300
DB 846 CAGAGCTGAG 905
QY 301 GluAlaLlthrsLlthrsProHisArgLysArgGlnGlyValAspProAlaSerPheHisPhe 320
DB 906 GAAGCCATCCACATCCACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
QY 321 LeuAspThrProLlthrsAlaLysValSerGluLeuGlnGlnArgLeuArgGlyTThlGluAsp 340
DB 1026 GGAAGCAGAGACTTGTGAGAGTCCCAAGGCCCAAGGAGAGAGAGAGAGAGAGAG 1085
QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaLlthrsProAlaValSerPro 380
DB 1086 GTGGCCAG 1145
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
DB 1146 TCCGCCCCCTGGTGCACCGCGCCCTCTCCCTCCCTAGAGAGAGAGAGAGAGAGAG 1205
QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
DB 1206 AAGCAG 1265
QY 421 AlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuValVal 440
DB 1266 GCTTCAAGTGGCCCTGTCTGTGGGCGGAGACACTGTGGGAGAGTGGCCGCTTGTGTG 1325
QY 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluLlthrsHisLlthrsGluHis 460
DB 1326 TATGAGAGCAG 1385
QY 461 HisHisHisTrpHisHisPheTyrGlnThr 470

Db 1386 CACCACCATTCACCACTTCTACCAAGACA 1415

|||||

RESULT 4
BC051288
LOCUS
DEFINITION Homo sapiens naked cuticle homolog 1 (Drosophila), mRNA (cDNA clone MGC:59666 IMAGE:6646089), complete cds.
ACCESSION BC051288
VERSION BC051288.1 GI:30410965
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2604)
Straussberg R.D., Collins E.A., Grouse L.H., Dege J.G., Krausner R.D., Colling F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueding T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A., Whiting M., Madan A., Young A.C., Shvedchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalusz D.E., Scherch A., Schein J.E., Jones S.J. and Marz M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)
12477932
2 (bases 1 to 2604)
Straussberg R.
Direct Submission
Submitted (25-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Akhter N., Ayele K., Beckstrom-Sternberg S.M., Benjamin B., Blakeley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S., Dietrich N.L., Granite S., Guan X., Gupta J., Hsieh F., Hansen N., Ho S.-L., Karling B., Kwong P., Latic P., Lega R., Madu Q.L., Maitello C., Maekert B., Mastrian S.D., McCloskey J.C., McDowell J., Pearson R., Stancirpop S., Thomas P.J., Touchman J.W., Tsung A., Vogt J.T., Walker M.A., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and Green E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 47 Row: 0 Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20127660, Location/Qualifiers
1..2604
/organism="Homo sapiens"

gene

CDS

mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:59666 IMAGE:6646089"
/tissue_type="Ovary, teratocarcinoma"
/clone_id="NIH_MGC_109"
/lab_host="DH10B-R"
/note="Vector: pOT87"
1..2604
/gene="NKD1"
/db_xref="LocusID:85407"
/db_xref="MIM:607851"
225..1637
/gene="NKD1"
/codon_start=1
/product="naked cuticle homolog 1"
/protein_id="AAH51288.1"
/db_xref="GI:30410966"
/db_xref="LocusID:85407"
/db_xref="MIM:607851"

ORIGIN

Alignment Scores:
Pred. No: 6,35e-112 Length: 2604
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-993-966-7 (1-470) x BC051288 (1-2604)

QY 1 MetGlyLeuLeuHisSerLeuProAlaAlaValCysValArgArgGluSerProGluGly 20
Db 225 ATGGGAAACTTCACTCAAGCCGCGCGGTGGCAAGGCGAGAGCCGGAAGGT 284
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgGlyGly11GluGluTTP11GlyArg 40
Db 285 GACAGCTTCCCGCGAGCGCTGCTGGCTCGGAAGGCATCAAGAGTGGATGGGAGA 344
QY 41 GlnArgCysProGlyGlyVal1SerGlyProArgGluLeuArgLeuAlaGlyThr11Gly 60
Db 345 CAGGCGTCCCGCGGCGGTGTCTCGGAGCCCGACAGCTGGCGGTGGCGGACCATAGGC 404
QY 61 ArgSerThrArgGluLeuVal1GlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 405 CGAAGCACCCGGAGCGCTGTGGGAGAGTGTGAACAACGCTCAGCGAGAAAGAGAG 464
QY 81 AspAspPheArgLeuGluValAlaLeuProProGluTyrThrAspGlyLeuGlySerGly 100
Db 465 GACGACCTTGGCGTGAAGAGCGGCTGCTCGAAGAACTGACGGGCTGGGAGGGA 524
QY 101 AspGluTyrLeuSerGluVal1SerGluProCysProGlySerLeuValGluLeu 120
Db 525 GATGAGAAAGATGAGAGAGTGAAGAGCAAGCCCTGCCAGGCTCCAGAAAGCAAGTGAAG 584
QY 121 PheGluGluLeuGlnCysAspVal1SerMetGluGluAspSerArgGlnGluTTPThrPhe 140
Db 585 TTGGAAGAGCTTCAAGTGCAGCTGTCCATGAGAGAGACAGCGGACGAGTGAACCTTC 644
QY 141 ThrLeuTyrAspPheAspAsnAnglyVal1ThrArgGluAsp11ThrSerLeuLeu 160
Db 645 ACCGTGATGACTTGAACAACGCGGAGAGTCAACCGGAGAGCATCACAGCTTCTG 704
QY 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLeuMet 180

Db 705 CACACCTATATAGAGGTGGAGTCTCTGTCAACCACTCCCAACATCCAGAAATG 764
Qy 181 LeuArgValIysLeuThrValAlaProAspGlySerGlnSerIlyAspSerValLeuVal 200
Db 765 CTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGGACAAAGAGAGGTGCTTGTCTC 824
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlyProThrGluAspLeu 220
Db 825 AATCAGGCTACCTGACGACGCAAGGCCCGAGCAAGACCAAGCCACTGAGAGCTTG 884
Qy 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 885 CGAGAGCTGGAGAGAAAGACGACGAGCCCGCTCAGTTCAGGTTGACAGCCGCTGGAG 944
Qy 241 GlnSerGlyCysTrpThrIshIshCysValAspGluAsnIleGluArgArgAsnIleTrpLeu 260
Db 945 CAGTCTGGCTGTACCACTTGCGTAGATGAGAACATCGAGAGAGAAACCACTACTTA 1004
Qy 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db 1005 GATCTCCGCGAGATGAAACCTACACGTCATTTGGGCTGGCTCCCTTCGTTGAGCC 1064
Qy 281 GlnIlySerGluLeuProProArgTrpSerAspProThrArgSerArgSerIleGluPro 300
Db 1065 CAGAGCTCAGAACTGCCCCCGACCTCCATCCCATCGATTCGCTCCCATGAGACCG 1124
Qy 301 GlnAlaIleIshIshIleProIshArgIlyAspProGlnIlyValAspProAlaSerPheIshPhe 320
Db 1125 GAAGCCATCCACATCCACACCGAAAGCCCAAGGCGTGGACCGGCGCTCTTCCACTTC 1184
Qy 321 LeuAspThrProIleAlaIlySerGluLeuGlnIlyValAspGlyIshIshIshIshIshIsh 340
Db 1185 CTTGACACCCCAATCGCAGAGTTCAGAGTTCAGGACGAGCTCCGAGGACCCAGAGAC 1244
Qy 341 GlySerIlyshIshPheValArgSerProIlySerValIleGlySerValIleGlyIshIsh 360
Db 1245 GGGAGCAGACACTTTGTGAGGTCCCCCAAGGCCCAAGGACAAAGTGTGGTGGGACAC 1304
Qy 361 ValAlaArgGlyAlaArgAsnIlyAspProLeuGlyProAlaIleProAlaValAspPro 380
Db 1305 GTGGCCAGAGGGGCAAAACCAAGCCCTCTGGGACCGGCACTCCCTGCGTGTCCCC 1364
Qy 381 SerIleIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIsh 400
Db 1365 TCCGCCCACTGGCTGGCAGGCCCGGCTCTCTCCCTCAAGCCCTCGGACACAG 1424
Qy 401 LysIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIsh 420
Db 1425 AAGCACAAGACCGAGCCAGAGAGCAGCAGAGGCTGCCGGGCTGGCAGGACCACTG 1484
Qy 421 AlaSerGlyIlyProValLeuGlyValGluIshIshIshIshIshIshIshIshIshIsh 440
Db 1485 GCTTCAGGTGGCTCTGTCTGGGCGGAGCACTGGCGGAGTGGCCGCTTGTGGTGTG 1544
Qy 441 TyrGluSerGlnAlaGlyIshIshIshIshIshIshIshIshIshIshIshIshIshIsh 460
Db 1545 TATAGAGCCAGGCGGAGCAGCCGCTCCAGAGACATGAGCACACACACATGAAACAT 1604
Qy 461 HisIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIsh 470
Db 1605 CACCACCATTCACCACTTCTACAGACA 1634
RESULT 5
AR405917 1438 bp DNA linear PAT 18-DEC-2003
LOCUS AR405917
DEFINITION Sequence 5 from patent US 6630323.
ACCESSION AR405917
VERSION AR405917.1 GI:40154965
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 1438)
AUTHORS Scott, M., Zeng, W. and Wharton, K.
TITLE Naked cuticle genes and their uses
JOURNAL Patent: US 6630323-A 5 07-OCT-2003;
FEATURES
source location/Qualifiers
1..1438
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 6,17e-112 Length: 1438
Score: 2493.00 Matches: 469
Percent Similarity: 99.79% Conservative: 0
Best Local Similarity: 99.79% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: Gaps: 0
US-09-993-966-7 (1-470) x AR405917 (1-1438)
Qy 1 MetGlyIlyLeuIshIshSerIlyAspProAlaAlaValCysIlyAspArgGlyIshSerProGluGly 20
Db 8 ATGGGAAACCTTCACTCCAGACCGGCGCGGTGTGCAAGGCGCAGAGACCCGGAAGT 67
Qy 21 AspSerPheAlaValSerAlaAlaTrpAlaArgIlyIleGluGluTrpIleGlyArg 40
Db 68 GACAGCTTCCGCTGAGCGCTGCTGGCTCGGAAGGCAATCGAGAGTGGATCGGAGA 127
Qy 41 GlnArgCysProGlyIlyValIserGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db 128 CAGGCTGCGCGGCGGTGTCTCGGAGCCCGACAGCTGCGTGGCGGCGCACATGAGC 187
Qy 61 ArgSerThrArgGluLeuValIleAspValLeuAspAspThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCAGGAGCTGTGGGCGAGCTGTGAGACACGCTCAGCGAGAAAGAGAG 247
Qy 81 AspAspPheArgLeuGluValAlaLeuProProGluIlyshIshIshIshIshIshIshIsh 100
Db 248 GACGACTTGGCTGGAAGTGGCTTCTCTGAGAAAGCTAGCGGGCTGGGACAGGA 307
Qy 101 AspGluIlyAlaMetGluArgValIserGluProCysProGlySerIlySerIlyGlnLeuIly 120
Db 308 GATGAGAAAGATGAGAGAGTACGAGAACCTCGCCAGGCTCCAGAGAACAGCTGAG 367
Qy 121 PheGluGluLeuGlnCysAspValIserMetGluGluAspSerArgGlnIlyTrpThrPhe 140
Db 368 TTTGAAAGCTCCAGTGCAGCTGTCCATGAGAGGACAGCCGCGAGGTGACCTTC 427
Qy 141 ThrLeuTyrAspPheAspAsnAsnGlyIlyValThrArgGluAspIleThrSerLeuLeu 160
Db 428 ACCCTGATGACTTTGACCAACCAAGCAGAGTCACTCCGAGAGACATCACACGCTTCTG 487
Qy 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerIlyMet 180
Db 488 CACACCATATGAGGTGGTGACTCTCTGTCAACCACTCCCAATCCAGCAAGATG 547
Qy 181 LeuArgValIlyLeuThrValAlaProAspGlySerGlnSerIlyAspSerValLeuVal 200
Db 548 CTGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGCCAGCAAGAGAGAGGTCTTGTCTC 607
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlyProThrGluAspLeu 220
Db 608 AATCAGGCTACCTGACGACGCAAGGCCCGAGCAAGACCAAGCCCACTGAGAGCTG 667
Qy 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGAGAGCTGGAGAAAGACGACGAGCCCGCTCAGTTCAGGTTGACAGCCGCTGGAG 727
Qy 241 GlnSerGlyCysTrpThrIshIshCysValAspGluAsnIleGluArgArgAsnIleTrpLeu 260
Db 728 CAGTCTGGCTGTACCACTTGCGTAGATGAGAACATCGAGAGAGAAACCACTACTTA 787
Qy 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280

```

Db      788 GATTCGCCCGGATGAAAATACACCTCCCAATTTGGGCTGGCTCCCTTCCTGGGCC 847
Qy      281 GlnlyserGlnleuProProArgThrSerAnbProThraSerArgSerHieGluPro 300
Db      848 CAGAGTCAGAACTGCCCCCGGACCTTCAATCCCATCGATCTCGCTCCCATGAGCGG 907
Qy      301 GlnAlaIleHieGlieProHieArgLyseProGlnGlyValaAppProAlaSerPheHiePhe 320
Db      908 GAAGCCATCCACATCCACACCGGAAGCCCGCAGCGCTGAGCCGCGCTTCCTCCACTTC 967
Qy      321 LeuAspThrProIleAlaIleValaIseGlnleuGlnIleArgLeuArgLyseValaIleGlyAsp 340
Db      968 CTGGACACCCCAATCGCCAGAGTTCAGAGTTCAGACACCGCTCCGGGCACTCAGGAC 1027
Qy      341 GlySerLyseHiePheValaIseSerProLyseAlaGlnGlyLyseValaIleGlyValaIleHie 360
Db      1028 GGGAGCAAGACCTTGTGAGGTGCCCAAGGCCCAAGGCAAGAGTGTGGTGGCCAC 1087
Qy      361 ValAlaArgGlyAlaIleArgAnlyseProProleuGlyProAlaIleProAlaValaIsePro 380
Db      1088 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGAGCCCGCCATCCCTGCGGTGTCCTCC 1147
Qy      381 SerAlaHieLeuAlaIseSerProAlaLeuLeuProSerLeuAlaProleuGlyHieIle 400
Db      1148 TCGGCCACCTGGTGGCCAGCCCGCCCTCTCCCTCTAGCCCTCCGGGCAAG 1207
Qy      401 LyseHieLyseHieArgAlaIleGlyIseGlnGlnGlyCysArgGlyLeuGlnAlaProleu 420
Db      1208 AACCAACAGACCGAGCCAGAGGCAAGAGGAGGAGGCTGGGGGCTGGAGGCAACCTG 1267
Qy      421 AlseSerGlyGlyProValleuGlyArgGlyHieLeuArgGlyLeuProAlaLeuValaIle 440
Db      1268 GCCTCAGTGGCTGCTGCTGGGGGAGACCTGCGGAGGCTGCGCGCTGGTGTG 1327
Qy      441 TyGlySerGlnAlaGlyIseGlnProValaIleArgHieGlnHieHieHieHieHieHie 460
Db      1328 TATGAGAGCCAGGCGGCGGAGCGGCTCCAGAGATGAGCACACCAACCATGAACAT 1387
Qy      461 HieHieHieHieHieHieHieHieHieHieHieHieHieHieHieHieHieHieHie 470
Db      1388 CACCACCATTAACCACTTCTTACAGACA 1417

RESULT 6
AF343352      1416 bp      mRNA      linear      ROD 29-MAR-2001
LOCUS      AF343352
DEFINITION      Mus musculus Nkd mRNA, complete cds.
ACCESSION      AF343352
VERSION      AF343352.1
KEYWORDS      GI:13487304
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      1. (bases 1 to 1416)
AUTHORS      Yan, D., Wallingford, J.B., Sun, T.-Q., Nelson, A.M., Sakanaka, C.,
TITLE      Reinhard, C., Harland, R.M., Pantl, W.J. and Williams, L.T.
JOURNAL      Cell autonomous regulation of multiple Dishevelled-dependent
SUBMITT     pathways by mammalian Nkd
PUBMED      Proc. Natl. Acad. Sci. U.S.A. 98 (7), 3802-3807 (2001)
MEDLINE      21173639
PUBMED      11274398
PUBMED      2. (bases 1 to 1416)
AUTHORS      Yan, D., Wallingford, J.B., Sun, T.-Q., Nelson, A.M., Sakanaka, C.,
TITLE      Reinhard, C., Harland, R.M., Pantl, W.J. and Williams, L.T.
JOURNAL      Direct Submission
SUBMITT     Submitted (26-JAN-2001) Research, Chiron Corporation, 4560 Horton
PUBMED      Street, Emeryville, CA 94608, USA
FEATURES
source
Location/Qualifiers
1. 1416
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

```

```

CDS
1. 1416
/note="Nkd; similar to Drosophila naked cuticle"
/codon_start=1
/product="Nkd"
/protein_id="AAK27485.1"
/db_xref="GI:13487305"
/translation="MGKLSKPAAVCKRRESPEGDSPAVSAAMARKIEEWIGRQCP
GSVSGPQLAGTGVKGTRELVDTSREALGEBDEDFLEVALPEPKIDSLGSDS
KRMERLSEPGQASKQDLKFEELQDVSVEEDSQEWFTLYEDNNKVKREDITSL
HTIYEVDVSVNHSPTSSKTLRYALTVAPGSGSQSKSVLPNNHDTLOSTRPADTPAP
ELRGMEKKORAPLRFQDSDHLEQDCHVDEINERNNHYDLAIGENTYSGGPS
PSVAQSELPSPRIISNTPRSRSHPEAHITPRPGQVDPGSPFLIDTPPKASFLOR
LRGTQDGSKRFPVSPKRAQSKMKMGHCAKARAKRPELVPTHTYVSPNHLATSPALLP
TLAPLGHKHKHRAKSDASCRGLQGLAAGSTVMGREQVRELPAVVYSQAGAV
ORHHHHHHHHHHHYP"

ALIGNMENT Scores:
Pred. No.: 1.95e-97 Length: 1416
Score: 2192.50 Matches: 410
Percent Similarity: 91.06% Conservative: 18
Best Local Similarity: 87.23% Mismatches: 41
Query Match: 87.74% Indels: 1
DB: 10 Gaps: 1

US-09-993-966-7 (1-470) x AF343352 (1-1416)

Qy      1 MecGlyLyseLeuHieSerLyseProAlaAlaValCysLyseArgArgLyseProGlnGly 20
Db      1 ATGGGGAACCTTCACTGAGAGCGGGCGCGCTGTGCAAGGCGAGGAGCGCGGAAGGT 60
Qy      21 AspSerPheAlaValaIseAlaAlaTPAlaArgLyseGlyTleGlnGluTTPtleGlyArg 40
Db      61 GACAGCTTCTCTGTAACCGCTGCTGGGCAAGGAAGACATCGAGGTGATCGGAGG 120
Qy      41 GlArgCysProGlyGlyValaIseGlyProArgGlnleuArgleuAlaGlyThrIleGly 60
Db      121 CAGGCTGTCCAGGACCGCTCCAGACCCCGTCAAGTGAATGGAGGACACTGTGGT 180
Qy      61 ArgSerThrArgGlnleuValaIleGlyAspValaLeuArgHiePheHieSerGlnGlnGlu 80
Db      181 CGAGGCACTCGGGAACCTCGTGGGTGACCTTCAAGAGGCTCTCGGTGAGAGAGAG 240
Qy      81 AspAspPheArgLeuGlnValaIleuProProGlnIleThrAspGlyLeuGlySerGly 100
Db      241 GACGACTTCCCTCAGAGATGGCGCCCTGAGAGAAAGATCAACAGCTAAGTAGTGA 300
Qy      101 AspGlnlyseIleMetGlnArgValaIseGlnProCysProIleSerLyseIleVala 120
Db      301 GATGAGAGAGAAATGAGAGACTGAGCAACTGGCCAGGCTTCCAAAGAGCAAGCTAAG 360
Qy      121 PheGlnGlnleuGlnCysAspValaIseMetGlnGlnAspSerArgGlnGlnIleThrPhe 140
Db      361 TTGGAAGAGCTACAGTGTATGTCTGTGTGAGAGAGACAGCCGCGCAAGAGTGACTTTC 420
Qy      141 ThrleuTyAspPheAspAnaanglyLyseValaIleThrArgGlnAspIleThrSerleu 160
Db      421 ACTCTATATGACTTCGCAACAATGCAAGTGAACCGGTGAGGACATTACAGCTTGCTG 480
Qy      161 HieThrIleTyArgIleValaIleAspSerValaAspHieSerProThrSerSerIleMet 180
Db      481 CATACCATCTATAGAGTGTGACTCTCTGTGAACCATTTCCCAATCAAGCAAGACA 540
Qy      181 LeuArgValaIleSerThrValaIleProAspGlySerGlnSerLyseArgSerValleuVal 200
Db      541 CTGGGGTGAAGCTCACCGTGGCTCTGACGGAGCGAGGTAAAGAGAGAGCTCTTTC 600
Qy      201 ArgGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnIleThrLyseProThrGlnAspLeu 220
Db      601 AACCATACCATCTGAGAGCAACAGGCCCGGAGCAACCAAAACCCGCTGAGAGCTG 660
Qy      221 ArgSerTrogIleLyseIleArgAlaProleuArgPheGlnIleAspSerArgLeuGlu 240

```

```

Db      661  CGTGGCTGGAGAGAAGACGACGACCCCACTCAGTTCCAGGTGACAGCCACTTGAG 720
Qy      241  GlnSerGlyCysTyrHisGysValAspGluAsn111eglUarGArGAsnHisTyrLeu 260
Db      721  CACGCAAGACTGTACCAACCACTTGGTGTGATGAGAACATTGAGAGGAAACCACTACTCTA 780
Qy      261  AspleuAaGly111egiAaenYrThSerGlnpHeGlyProGlySerProSerVal1A 280
Db      781  GACTGGCGGGAGTGAAGAACTACACCTCTCAGTTTGACCGGGATCCCTTCGGTGGCC 840
Qy      281  GlnLysSerGluLeuProProArgThrSerAspProThrArgSerArgSerHisGluPro 300
Db      841  CAGAGTCAGAGGCTGCCCTCCGATCTCCAAACCCCACTGGCTCTCGCTCCCAAGAGCCA 900
Qy      301  GluAla111111111111111111111111111111111111111111111111111 320
Db      901  GAAGCTGCCCAACATCCACACCGGAGGCCCAAGAGTGTGACCCAGAGCTCTCCACACTC 960
Qy      321  LeuAspThrPro111111111111111111111111111111111111111111111 340
Db      961  CTTGACACCCCATTTTGCACAGGACATCAGAGCTCCAGCAAGCGCTCCGAGGACCTCAGAT 1020
Qy      341  GlySerLysHisPheValArgSerProLysValGlnGlyLysSerValGlyValGlyHis 360
Db      1021  GGGAGCAAGACATTTTGTGAGGTGCCCAAGGCCCAAGGCAAGACATGGGTATGGGCGAC 1080
Qy      361  ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAla1111111111111111 380
Db      1081  GGGGGCCAGAGGTGTGACAGAGCAAGAGCTCTCAGTACCCACACCCATCTCTCCCTCC 1140
Qy      381  SerAlaHisLeuAla1111111111111111111111111111111111111111111 400
Db      1141  TCTGCCCATCTGGCCACAGCCCAAGCCCTTCTCCCACTGTGACACCCCTGGGGCAAG 1200
Qy      401  LysHisLysHisArgAlaLysGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      1201  AAACACACAGCATGACGCAAGAGAGCAAGCGAGCTGCCGGGCTGTGACGAGGCCCTCTG 1260
Qy      421  AlaSerGlyLys---ProValLeuGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 439
Db      1261  GCGGCAAGAGAGCTCCACCGTCAATGGGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy      440  ValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisGln 459
Db      1321  GTGTACAGAGCCAGCGCTGGGCAAGCCGCTCCAGAGACCAAGACCATCACCAACAGAA 1380
Qy      460  HisHisHisHisTyrHisHisHisHisHisHisHisHisHisHisHisHisHisHis 469
Db      1381  CATCACCAACATATATCAACATCTTATCAG 1410

RESULT 7      1765 bp      mRNA      linear      ROD 29-JUN-2004
BC034838      MGC:40995 IMAGE:1429943, complete cds.
LOCUS      MGC:40995 IMAGE:1429943, complete cds.
DEFINITION      MGC:40995 IMAGE:1429943, complete cds.
ACCESSION      BC034838
VERSION      BC034838.1 GI:22028144
KEYWORDS      MGC.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1765)
Strausberg,R.L., Feltingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenman,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buettow,K.H., Schaefer,C.F., Bhac,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,W., Soares,M.B., Bonaldo,M.F., Casavant,T.D.,
Schnee,T.E., Brownstein,M.J., Udén,T.B., Yoshiyuki,S.,
Chen,H., Prange,C., Raha,S.S., Loguercio,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

```

```

TITLE      JOURNAL
PUBMED     12477932
REFERENCE  2 (bases 1 to 1765)
AUTHORS    Strausberg,R.
TITLE      Direct Submission
JOURNAL     Submitted (29-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapsb-remail.nih.gov
            Tissue Procurement: Marcello Bento Soares, Ph.D.
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            Info@cgsc.bc.ca
            Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
            Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
            Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
            Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
            Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lisa Prehbu,
            Parvaneh Saeedi, Dr Santos, Angélique Scherch, Ursula Skalske,
            Duane Smalls, Jeff Scott, Miranda Tsai, George Yang, Jacquié
            Schein, Asim Siddiqui, Rob Holt, Marco Marra.

FEATURES             source
CDS
    gene
        1..1765
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="MGC:40995 IMAGE:1429943"
            /tissue_type="Mammary gland"
            /clone_11b="Soares_mammary_gland_NbMWG"
            /lab_host="DH10B"
            /note="Vector: pT7T3-Pac"
            1..1765
            /gene="Nkd1"
            /db_xref="LOCUSID:93960"
            /db_xref="MGI:2135954"
            138..1553
            /gene="Nkd1"
            /codon_start=1
            /product="naked cuticle 1 homolog"
            /protein_id="AAH34838.1"
            /db_xref="GI:22028145"
            /db_xref="LOCUSID:93960"
            /db_xref="MGI:2135954"
            /translacion="MGKLSKPAVCKRRESPEGDSFVSAAMARKIEBMIGRQCP
            KSVSPGROLAIVGKGTRELVGDSRELAIGEDDDPLFVALPEKIDLSGDB
            KMERPLSEPGQASKKOLKEELOCVDVSVEEDSOEWFTLYDPNNKXVRBDTSL
            HTIYVYDSSVNHSPSSKTLRYKLYTAPGSSGSKSVLNHNDIOSTRPRDKPAB
            ELGWKSEKQAPLRFQDSHLEQDCTHNCVDENIERNNYLDLAGIENTTSQGPDS
            PSVAQKSELPRIISNPRISHPEDAAHVHRPQGVDPDSFILDTFPAVASLQOR
            LRITQDSKIFVSPKAOQKMGMGHARGARBPPLVPTHTVSPAHNATSPALLP
            TLAFLGHKKHKBKESQASCRGIGPLAAGSGTVMGRQVRELPAVVYVESQAGQAV

```

ORHHHHHHHHHHHHFYQP"

Alignment Scores:

Pred. No.: 2,81e-97 Length: 1765
 Score: 2191.50 Matches: 409
 Percent Similarity: 91.06% Conservative: 19
 Best Local Similarity: 87.02% Mismatches: 41
 Query Match: 87.70% Indels: 1
 DB: 10 Gaps: 1

US-09-993-966-7 (1-470) x BC034838 (1-1765)

QY 1 MetGlyLeuHsSerLysProAlaAlaValCysLysArgArgLysSerProGluGly 20
 Db 138 ATGGGGAATCTCACTGAGAGCCGGCGCGCTGTGCAAGCGCAGAGAGCCCGGAAGGT 197
 QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyIleGluGluTTPAlaGlyArg 40
 Db 198 GACAGCTTCTGTAAACGCTGCTGGGCAAGAAAGCATCGAGAGTGGATCGGAGG 257
 QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
 Db 258 CACGCTGTCCAGGACGCTCTCAGAGACCCGCTCAGCTGAGATTGGCAGGCACTTGGT 317
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
 Db 318 CGAGGCACTGGGAACTGTGGGTGACACTTCTAGAGAGCTCTCGGTGAGAGAGAG 377
 QY 81 AspAspPheArgLeuGluValAlaLeuProProGluThrThrAspGlyLeuGlySerGly 100
 Db 378 GAGAGCTTCCCTTAAGAGTGGCCCTGCGCTGAGAAAGATTGACACCTAGGTAGTGA 437
 QY 101 AspGluLysLysMetGluArgValSerGluProCysProGlySerLysGluLeuLys 120
 Db 438 GATGAGAGAGAGATGAGAGACTGAGGAGCACTGCGCAGGCTCCAGAGAGAGCTCAG 497
 QY 121 PheGluGluLeuGlnCysAspValSerMetGluLysSerArgGlnGluTTPThrPhe 140
 Db 498 TTTGAAAGAGTGAAGTGTGATGTCTGTGAGAGAGAGAGCCGCAAGAGTGAAGTTC 557
 QY 141 ThrLeuThrAspPheAspAsnGlyLysValThrArgLysAspIleThrSerLeuLeu 160
 Db 558 ACTCTATATATCTTCCAAACAATGSCAAAGTGACCCGTGAGACATTACAGCTTCTG 617
 QY 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180
 Db 618 CATACCATCTATGAAAGTGTGACTCTGTGAAACATTCCCAATCAGCAAGACACA 677
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
 Db 678 CTGCGGTGAGAGTCAACCGGTCTCCGAGCGGAGCAAGTGAAGAGAGAGCTCTTTC 737
 QY 201 AsnGluAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
 Db 738 AACCATACCATCTGCAAGACACAGCCCGCAGACAGACACCAACCCGCTGAGAGCTG 797
 QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 Db 798 CGTGGCTGGGAGAGAGAGAGAGAGCGGAGCCCACTCAGGTTCCAGAGTGCACACCTGAG 857
 QY 241 GlnSerGlyCysTyrHisGlyValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
 Db 858 CAGCCAGACAGCTACACCACTTCGTGAGTGAAGAACTTGAAGAGAGAAACCACTACTA 917
 QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 Db 918 GACCTGGCGGAGAGAGAGAGAGAGAGTCTCAGATTGACCGGAGATCCCTTCGGTGGCC 977
 QY 281 GlnLysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
 Db 978 CAGAAAGTCAGAGAGTGGCCCTCGAATCTCAACCCCACTGCTCTGCTCCACAGAGCA 1037

QY 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
 Db 1038 GAACTGCCACAGCTCCACACCGAGGCCCCAAGAGTGGACCCAGGCTCTTCCACTC 1097
 QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgLysThrGlnAsp 340
 Db 1098 CTGACACCCCATTTGGCCAAAGGATCAGAGCTCCAGCAAGGCTCCGGGAGCAAGAT 1157
 QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHis 360
 Db 1158 GGGAGCAAGCACTTTGTGAGAGTCCCCCAAGGCCCAAGAGCAAGAACTAGGTATGGGCA 1217
 QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 1218 GGGGACAGAGTGCAGAAAGACAGCTTCACTGTATCCACCACTACTGTCTCCCC 1277
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuAlaHis 400
 Db 1278 TGTGCCATCTGGCCACACAGCCCAAGCCCTTCTCCCACTGGGACCCCTGGGACAG 1337
 QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 Db 1338 AAACACAAAGCATGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1397
 QY 421 AlaSerGlyGly---ProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 439
 Db 1398 GCTCAGAGAGAGCTCCACAGTATGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1457
 QY 440 ValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisGlu 459
 Db 1458 GTGTACAG 1517
 QY 460 HisHisHisHisTyrHisHisPheTyrGln 469
 Db 1518 CATCACCACTTACCACTTCTATCAG 1547
 RESULT 8
 AF358134
 LOCUS AF358134 1671 bp mRNA linear ROD 29-MAY-2001
 DEFINITION Mus musculus naked cuticle-1 (Nkd1) mRNA, complete cds.
 ACCESSION AF358134
 VERSION AF358134.1 GI:14211711
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1671)
 Wharton, K.A., Jr., Zimmermann, G., Rouselet, R. and Scott, M.P. Vertebrate proteins related to Drosophila Naked Cuticle bind Dishevelled and antagonize Wnt signaling Dev. Biol. 234 (1), 93-106 (2001)
 JOURNAL MEDLINE 2125530
 PUBMED 11356022
 REFERENCE 2 (bases 1 to 1671)
 Wharton, K.A., Jr., Zimmermann, G. and Scott, M.P. Direct Submision Submitted (08-MAR-2001) Pathology and Molecular Biology, University of Texas Southwestern Medical School, 5323 Harry Hines Blvd., Dallas, TX 75390-9072, USA
 FEATURES
 source
 1.1671
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="L84"
 /sex="female"
 /tissue_type="lung"
 /dev_stage="6-8 weeks"
 1.1671
 /gene="Nkd1"
 80..1495
 CDS

```

/gene="Nkd1"
/note="contains EFX domain, similar to Drosophila
melanogaster naked cuticle (nkd)"
/codon_start=1
/product="naked cuticle-1"
/protein_id="AK57483.1"
/db_xref="GI:14211712"
/translation="MGKLSKPAVCKRSPSPGSDPAVSAAMKGIIEIGRORCP
GSVSGRQLRLACTVGKRTREIVGDSRELAGEDEDFPLEVALPREKIDISGSDP
KRMRLSEPGQASKQDKFEELQDVSVESDSROEMPTLYDPNNKVTREDITSLD
HTIYEVDSVSNHPTSSKTLRYVLIVAPGDSQSKSVLFNHTDQSTRPADTSPAE
ELRGEKKQAPLRFQGDSHLEQDCEYHCEVDENIERHNYLDLAGIENTYSGFQPS
PSVAQSELPRIISNPTSRSHPEAAHPIHRRGVDPGSPFLDTPFAKASELQOR
LRCTQDSKHFVSPRAQGNMGMGHGAARABSKPLVPTHTYSPSAHLATSPALLP
TLAPLGHKHKRAKESQASCRGIQSPILAAGSSTWGREQVRELPAVMYTESQAGAV
ORHEHHHHHHHHHYHPTOP"

```

ORIGIN

Alignment Scores:

Pred. No.:	3.3e-97	Length:	1671
Score:	2189.50	Matches:	409
Percent Similarity:	91.06%	Conservative:	19
Best Local Similarity:	87.02%	Mismatches:	41
Query Match:	87.62%	Indels:	1
DB:	10	Gaps:	1

US-09-993-966-7 (1-470) x AF358134 (1-1671)

```

QY 1 MetGlyLysLeuHisSerIysProAlaAlaValCylsYsArgArgIuserProGluGly 20
Db 80 ATGGGGAACCTTCACCTCGAAGCCGCGCGCTGTGCAAGCGCAGAGAGAGAGAGAGGT 139
QY 21 AspSerTherAlaValSerAlaAlaTTPAlaArgIysArgIyleGluGluTTPILegIYArg 40
Db 140 GACAGCTTTCTGTGACGCTGTGTCGCAAGAAAGCATCGAGAGTGGATCGGAGG 199
QY 41 GlArgCysProGlyGlyValSerGlyProArgGluInleuArgIysLeuAlaGlyThrIleGly 60
Db 200 CAGCGCTGTCCAGGACGCTTCAGAGACCCCGTACGTGAGATTGGACGACCTGTGGT 259
QY 61 ArgSerTherArgIleuValGlyAspValLeuArgAspThrIleuSerGluGluGlu 80
Db 260 CGAGGCACTCGGGAACCTCGTGGGTGACACTTCAGAGAGGCTCTCGGTGAGAGAGAGAG 319
QY 81 AspAspPheArgIleuGluValAlaLeuProGluIlyThrArgGlyLeuGlySerGly 100
Db 320 GACGACCTTCCCTTGAAGAGTGGCCCTGCGCTGAGAAAGATCGACACCTAGGTAGTGA 379
QY 101 AspGluLysLysMetGluArgValSerGluProCysArgProGlySerIysLysGluInleu 120
Db 380 GATGAGAGAGATGAGAGACTGAGGACCTGCGCAGGCTCCAGAGAGAGAGAGAGAG 439
QY 121 PheGluGluLeuGluCysAspValSerMetGluGluAspSerArgIleuTTPThrPhe 140
Db 440 TTTCGAAGCTACAGTGTGATGTCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 499
QY 141 ThrLeuTrpAspPheAspAspAspArgIlyValThrArgIleuAspIleThrSerLeuLeu 160
Db 500 ACTCTAATATGACTTCGACCAACATGCAAGTGAACCGGTGAGCAATTAACAGCTTGGT 559
QY 161 HisThrIleTrpGluValAlaAspSerSerValAsnHisSerProThrSerSerIysMet 180
Db 560 CATACCATCTATGAGAGTGTGACTCTCTGTGAAACATTTCCCAATCAAGAGAGAGACA 619
QY 181 LeuArgValIysLeuThrValAlaProAspGlySerGluSerIysArgSerValLeuVal 200
Db 620 CTGGGGGTGAAGCTCACCGTGGCTCTGACGGAGGAGGAGAGAGAGAGAGAGAGAG 679
QY 201 AsnGluAlaAspLeuGluInserAlaArgProArgAlaGluThrIlyProThrGluAspLeu 220
Db 680 AACCATATCCATCTGACAGAGCAAGAGCCCGAGACAGACCAAAACCCGCTGAGAGAGCTG 739
QY 221 ArgSerTrpGluLysLysGluArgAlaProleuArgPheGluGluAspSerArgLeuGlu 240

```

```

Db 740 CGTGGCTGGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 799
QY 241 GluSerGlyCysTrpThrHisCysValAspGluAsnIleGluLysArgAsnHisTrpLeu 260
Db 800 CACCCAGACTGCTTACCAACCAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGluPheGlyProGlySerProSerValAla 280
Db 860 GACCTGGCGGAGATAGAGAACTACAGCTCTCACTTTGGACCGGAGATCCCTTCGTGGCC 919
QY 281 GluLysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 920 CAGAGTCAAGAGCTGCCCTCCCAATCTCAACCCCACTGCTCGCTCCCAAGAGCCA 979
QY 301 GluAlaIleHisIleProHisArgIlyProGluGlyValAspProAlaSerPheHisPhe 320
Db 980 GAACTGCCCACTCCACACCGAGAGGCCCAAGAGTGTGAGCCAGGCTCTTCACCTC 1039
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGluGluArgLeuArgGlyThrGluAsp 340
Db 1040 CTTGACACCCCAATTTGCAAGGATCAAGACTTCAGACAGCGCTCCGGGAGACTCAGAGAT 1099
QY 341 GlySerLysHisPheValArgSerProLysAlaGluGlyLysSerValGlyValHis 360
Db 1100 GGGAGAGAGACATTTGTGAGGTCCCCCAAGGCCCAAGGCAAGAAACATGGATAGGCGCAC 1159
QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1160 GGGGCGCAGAGTCCAGAGAACCAAGCCCTCCACTGTATCCCAACCCCACTACTGTTCCCCC 1219
QY 381 SerAlaHisLeuAlaAspProAlaLeuLeuProSerLeuAlaProleuAlaPheLys 400
Db 1220 TCTGCCATCTGGCCACAGCCCAAGCCCTTCTCCCACTGGCAGCCCTCGGGCACAAG 1279
QY 401 LysHisLysHisArgAlaLysGluSerGluGluGlyCysArgGlyLeuGluAlaProleu 420
Db 1280 AAACACAGCATGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
QY 421 AlaSerGlyLys--ProValLeuGlyArgGluHisLysLeuArgIleuProAlaLeuVal 439
Db 1340 GCTGCAGAGGCTCCACCGCTCATGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1399
QY 440 ValTrpGluSerGluAlaGlyGluProValGluArgHisGluHisHisHisHisGlu 459
Db 1400 GTTGACAGAGCCAGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459
QY 460 HisHisHisHisTrpHisHisPheTrpGlu 469
Db 1460 CATCACCATATATCAACACTTCTATCAG 1489

```

RESULT 9

AR405916

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:	1.82e-96	Length:	1731
------------	----------	---------	------

```

1.1731
/mol_type="genomic DNA"

```

```

1 (bases 1 to 1731)
Scott, M., Zeng, W. and Wharton, K.
Naked cuticle genes and their uses
Patent: US 6630323-A 3 07-Oct-2003;
Location/Qualifiers
1.1731

```

Score:	2174.50	Matches:	407
Percent Similarity:	90.64%	Conservative:	19
Best Local Similarity:	86.60%	Mismatches:	43
Query Match:	87.01%	Indels:	1
DB:	6	Gaps:	1
US-09-993-966-7 (1-470) x AR405916 (1-1731)			
QY	1	MeGlyLysLeuHisSerLysProAlaIaValaCysAlaYsaRgaRgiuSerProGluGly	20
DB	140	ATGGGGAACTTCACTCGAAGCCGGCCGGTGTGCAAGCCAGAGAGAGCCGGAAAGT	199
QY	21	AspSerPheAlaValaSerAlaIaTropAlaArgLysLeuGluTyrPheGlyArg	40
DB	200	GACAGCTTGTCTGAAGCGCTGTGGCAAGAAAGCATCGAGAGTGATCGGAGG	259
QY	41	GlnArgCysProGlyGlyValaSerGlyProArgGluLeuArgLysLeuGluTyrPheGly	60
DB	260	CACGGCTGTCAAGGACGCTCTCAAGACCCTGTACGTGAGATTGGCAGGCACTGTGCT	319
QY	61	ArgSerThrArgGluLeuValaGlyAspValaLeuArgAspThrLeuSerGluGluGlu	80
DB	320	CGAGGCACTCGGAACTCGTGGGTGACATTCCTAGAGAGCTCTCGGTGAGAGAGAG	379
QY	81	AspAspPheArgLeuGluValaIaLeuProProGluLysThrAspGlyLeuGlySerGly	100
DB	380	GAGGACTTCCCCCTAGAAAGTGGCCCTGCTGAGAAAGATCGACAGCTAGGTAGTGA	439
QY	101	AspGluLysLysMetGluArgValaSerGluProCysProGlySerLysLysGluLeuLys	120
DB	440	GATGAGAAAGATGAGAGACCTGAGGAACTGGCCAGGCTTCAAGAAAGCACTAG	499
QY	121	PheGluGluLeuGluCysAspValaSerMetGluGluAspSerArgGluTyrPhe	140
DB	500	TTTGAAGAGCTACAGTGTGATGTCTCTGTGAGAGAGACAGCGGCAAGTGTGACTTTC	559
QY	141	ThrLeuTyrAspPheAspAsnGlyLysValaThrArgGluAspLysLeuSerLeuLeu	160
DB	560	ACTCTATATGACTTCCACAAATGGAAGTACCCGTGAGAGACATTCAGAGTGTGCTG	619
QY	161	HisThrLysLysValaValaAspSerSerValaAsnHisSerProThrSerSerLysMet	180
DB	620	CATACCATCATTAAGGTGTTGACTCTCTGTGAACATTCCTCCACATCAAGCAAGACA	679
QY	181	LeuArgValaLysLeuThrValaIaProAspGlySerGluSerLysValaSerValaLeuVal	200
DB	680	CTCGGGGTGAAGCTCACCGTGGCTCTGACGGAGCCAGGTAAAGAGAGCGTCTTTC	729
QY	201	AsnGluAlaAspLeuGluSerLysAlaArgProArgAlaGluThrLysProThrGluAspLeu	220
DB	740	AACCATATCCATTTGACAGACACAAAGGCCCGGAGCAACAAACCCGTGAGAGCTG	799
QY	221	ArgSerTyrGluLysLysGluArgAlaProLeuArgPheGluGluAspSerArgLysGlu	240
DB	800	CTGGGCTGGAGAAAGAGAGAGAGCCCTCAAGGTTCCAGGGTGAACACCTGGAG	859
QY	241	GlnSerGlyCysTyrHisLysCysValaAspGluAsnLysLeuArgArgAsnHisTyrLeu	260
DB	860	CAGCCAGACTGTACACCATTCGTGTGATGAGAACTTGAAGAGAAACCACTACTA	919
QY	261	AspLeuAlaGlyLysLeuGluLeuTyrThrSerGluPheGlyProGlySerProSerValaIa	280
DB	920	GACCTGGCGGGATTAAGAACTACAGCTTCAGTTTGGACCGGAGTCCCTTCGGTGGCC	979
QY	281	GlnLysSerGluLeuProArgThrSerAsnProThrArgSerArgSerHisGluPro	300
DB	980	CAGAAATTCAGAGTGGCCCTTCAATCTCAAACTCCACTGCTCTGCTCCACAGACCA	1039
QY	301	GlnAlaLysHisLysProHisArgLysProGluGlyValaAspProAlaSerPheHisPhe	320
DB	1040	GAAAGTCCCATCCACACCGAGAGCCCAAGGTGTGAGACCAAGGCTCTTCCACCTC	1099
QY	321	LeuAspThrProLysValaSerGluLeuGluGlnArgLeuArgGlyThrGlnAsp	340

DB	1100	CTTGACACCCCATTTGCTCAAGGACATTCAGACTCCAGAACGGCTCCGGGCACTCAGAT	1159
QY	341	GlySerLysHisPheValaArgSerProLysAlaGlnGlyLysSerValaGlyValaGlyHis	360
DB	1160	GGAGAGAGCACTTTGTGAGGTGCCCAAGGCCCAAGAGAAAGATAGGTATGGGCAC	1219
QY	361	ValaIaArgGlyAlaArgAsnLysProProLeuGlyProAlaLysProAlaValaSerPro	380
DB	1220	GGGGCCAGAGGTGCAAGAACCAAGCTCCACTGTGTACCCACCACTTACTGTCTCCCC	1279
QY	381	SerAlaHisLeuAlaIaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys	400
DB	1280	TCTGCCATCTGGCCACACCGCCCTTCTCTCCCACTGGGCACTGGGGCACAAAG	1339
QY	401	LysHisLysHisArgAlaLysGluSerGluGlnGlyCysArgGlyLeuGlnAlaProLeu	420
DB	1340	AAACACAGCATGAGGCAAGAGAGAGCCAGGCTGCGGGGCTTGCAGGGCCCTG	1399
QY	421	AlaSerGlyGly--ProValaLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal	439
DB	1400	GCTGCAAGAGCTTCCACCTGTATGGGGCGGAGCAGGTGAGGAGCTGCTGCCGTATG	1459
QY	440	ValTyrGluSerGlnAlaGlyGlnProValaGlnArgHisGluHisHisHisGlu	459
DB	1460	GTGTACAGAGCCAGGCTAGGAGCCGTCCAGAGACAGAACCATTCACACCAAGAA	1519
QY	460	HisHisHisHisTyrHisHisPheTyrGln	469
DB	1520	CATCACCATTTATCACTTCTATCAG	1549
RESULT 10			
LOCUS	AX172351	1401 bp	DNA linear PAT 03-JUN-2001
DEFINITION	Sequence 1 from Patent WO0144279.		
ACCESSION	AX172351		
VERSION	AX172351.1	GI:14597503	
KEYWORDS			
SOURCE			
ORGANISM	Mus musculus (house mouse)		
REFERENCE			
AUTHORS	Van D. and Williams, L.T.		
TITLE	Mammalian disclosed-associated proteins		
JOURNAL	Patent: WO 0144279-A 1 21-JUN-2001;		
Chiron Corporation (US)			
FEATURES			
source	Location/Qualifiers		
	1..1401		
	/organism="Mus musculus"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:10090"		
ORIGIN			
Alignment Scores:			
Pred. No.:	3,09e-96	Length:	1401
Score:	2167.50	Matches:	406
Percent Similarity:	90.79%	Conservative:	18
Best Local Similarity:	86.94%	Mismatches:	42
Query Match:	86.73%	Indels:	1
DB:	6	Gaps:	1
US-09-993-966-7 (1-470) x AX172351 (1-1401)			
QY	1	MeGlyLysLeuHisSerLysProAlaIaValaCysAlaYsaRgaRgiuSerProGluGly	20
DB	1	ATGGGGAACTTCACTCGAAGCCGGCCGGTGTGCAAGCCAGAGAGAGCCGGAAAGT	60
QY	21	AspSerPheAlaValaSerAlaIaTropAlaArgLysLeuGluTyrPheGlyArg	40
DB	61	GACAGCTTGTCTGAAGCGCTGTGGCAAGAAAGCATCGAGAGTGATCGGAGG	120
QY	41	GlnArgCysProGlyGlyValaSerGlyProArgGluLeuArgLysLeuGluTyrPheGly	60


```

QY 207 SerAlaArgProAlaArgIuThyrLeuProThrcIuLeuLeuAlaGserTrrgluLyLys 226
DB 442 AGGCGAAGGCGCCGAGAGAGACCAACCACTGAGAGACCTGGAGAGAGAGAGAGAG 501
QY 227 GAlaArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlnGlnSerGlyCysTyrHis 246
DB 502 CACGAGAGCCCGCTCAGGCTTCAGAGGTGACAGCCGCTGAGAGAGCTGGCTGCTTACAC 561
QY 247 HisCysValaArgIuLeuIleGluArgArgAsnHisTrrLeuLeuAlaGlyIleGlu 266
DB 562 CATTGGCTGATGATGAGAACTGAGAGAGAGAAACCACTTAACTCTCGCCGGGATGAA 621
QY 267 AserTrrThrsGlnPheGlnPrrGlySerProSerAlaAlaGlnLysSerGlnLeuPro 286
DB 622 AACTACACGCTCCCAATTTGGGCTCGCTCCCTCCCTCCCTGAGAGAGAGAGAGAGAG 681
QY 287 ProArgThrsSerAsnProThrsArgSerArgSerHisGlnProGlnAlaIleHisIlePro 306
DB 682 CCCCCGACCTCCCAATCCCACTCGATCTCGCTCCCATAGCCGAGAGAGAGAGAGAGAG 741
QY 307 HisArgLysProGlnGlnValaArgProAlaSerPheHisPheLeuAspThrProIleAla 326
DB 742 CACCGAAGGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801
QY 327 LysValaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 346
DB 802 AAGGTCTCAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 861
QY 347 ArgSerProLysAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 366
DB 862 AGGTCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
QY 367 AsnLysPrrProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 386
DB 891 ----- 891
QY 387 SerProAlaLeuLeuProSerLeuAlaProLeuGlnGlnGlnGlnGlnGlnGlnGln 406
DB 892 -----GCCCGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927
QY 407 LysGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 426
DB 928 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 987
QY 427 LeuGlnArgGlnHisLeuArgGlnLeuProAlaLeuValValTrrGlnSerGlnAlaGly 446
DB 988 CTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1047
QY 447 GlnPro 448
DB 1048 CAGCCG 1053
RESULT 12
AC145175 207486 bp DNA linear HTG 15-AUG-2003
LOCUS Pan troglodytes clone CH251-32B1, WORKING DRAFT SEQUENCE, 11
DEFINITION
ordered pieces.
AC145175
AC145175.2 GI:33667124
VERSION HTG, HTGS PHASE2, HTGS_DRAFT.
KEYWORDS Pan troglodytes (chimpanzee)
SOURCE Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 207486)
Antoniellis A., Ayala K., Beckstrom-Stenberg S.M., Benjamin B.,
Blakesley R.W., Bouffard G.G., Brinkley C., Brooks S., Carls G.,
Chu G., Coleman B., Coleman H., Engle J., Granite S., Guan X.,
Gupte J., Haghighi P., Han J., Hansen N., Ho S.-D., Hu P.,
Hurdle B., Idol J.R., Karlins E., Kwong P., Latic P., Lee-Lin S.-Q.,
Legaspi R., Maduro Q. L., Maduro V.B., Margulies E.H., Masello C.,
Maskeri B., McDowell J., Paguirigan C., Pearson R., Portnov M.E.,
Praesed A., Reddix-Dugue N., Schandler K., Schueler M.G., Shah K.,

```

REFERENCE AUTHORS TITLE JOURNAL COMMENT

Sison, C., Stantirrop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
Vogt, J.L., Wetherby, K.D., Young, A. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 207486)
Green, E.D.
Direct Submission
Submitted (12-JUN-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
3 (bases 1 to 207486)
Green, E.D.
Direct Submission
Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
On Aug 15, 2003 this sequence version replaced gi:31623117.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: e91
Center clone name: 032B01

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 204881 bases at least Q40
Consensus quality: 205776 bases at least Q40
Consensus quality: 206263 bases at least Q20
Insert size: 192000; agarose-fp
Quality coverage: 12.59x in Q20 bases; agarose-fp
Quality coverage: 11.71x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
provided by the submitter.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
1
430: contig of 430 bp in length
431
530: gap of unknown length
20322
20421: contig of 19791 bp in length
20322
20421: gap of unknown length
20422
99662: contig of 79241 bp in length
99663
99762: gap of unknown length
99763
106538: contig of 6776 bp in length
106539
106638: gap of unknown length
106639
113489: contig of 6851 bp in length
113490
113589: gap of unknown length
113490
143460: contig of 29871 bp in length
143461
143560: gap of unknown length
143561
149849: contig of 6289 bp in length
149850
149949: gap of unknown length
149950
151973: contig of 2024 bp in length
151974
152073: gap of unknown length

* 152074 160497: contig of 8424 bp in length
* 160498 160597: gap of unknown length
* 160538 170482: contig of 9885 bp in length
* 170483 170582: gap of unknown length
* 170583 207486: contig of 36904 bp in length.
location/Qualifiers
1. 207486

source

/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone_1id="CH251-3281"
/clone_1id="CH251"
1. 430
misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
531. 20321
misc_feature
/note="assembly_fragment"
20422. 99662
/note="assembly_fragment"
96974. 207486
/note="clone overlaps with GenBank Accession Number
AC145238 clone CH251-568P19 (center project name esk)"
99763. 106538
misc_feature
/note="assembly_fragment"
106639. 113489
misc_feature
/note="assembly_fragment"
113590. 143460
misc_feature
/note="assembly_fragment"
143561. 149849
misc_feature
/note="assembly_fragment"
149950. 151973
misc_feature
/note="assembly_fragment"
152074. 160497
misc_feature
/note="assembly_fragment"
160598. 170482
misc_feature
/note="assembly_fragment"
170583. 207486
/note="assembly_fragment"
missing 77 clone end on 3' end of insert"

ORIGIN

Alignment Scores:

Pred. No.: 1 69e-47 Length: 207486
Score: 1210.50 Matches: 314
Percent Similarity: 26.33% Conservative: 2
Best Local Similarity: 26.17% Mismatches: 1
Query Match: 48.44% Indels: 883
DB: 2 Gaps: 3

US-09-993-966-7 (1-470) x AC145175 (1-207486)

QY 154 GluApIleThrSerLeuLeuHsrThrIleTygIuValValaPseSerValaShis 173
Db 177455 CAGAGATCAACAGCTTGTCGACACCATGTATGAGGTGGAGCTCCTGTCAACAC 177514
QY 174 SerProthrserserilysmetleuArqVallyleuthrValaIaProlaPglySerGln 193
Db 177515 TCCCAACATCCAGCAAGATGCTGGGGTAAAGCTCACCGTGGCCCCGATGGCAGC 177574
QY 194 SerlyArserValleuValaShngIaA----- 203
Db 177575 AGCAAGAGAGCGCTCTTGTCAATCAAGCTGTGTAAGGCTGCAAGGCTGAGCCTGGGAAG 177634
QY 203 ----- 203
Db 177635 CAATGCTCCTCCATCTCAGAAAGAACAGAGCTTCTTGGAAGGCGGAAAGGAACCTG 177694
QY 203 ----- 203
Db 177695 TGGGCTTTCTGGGGAGCTGCTTTGTGTGACAGGTGGGTTCMAATGGGTCTTGAAAGG 177754
QY 203 ----- 203

Db 177755 TAGAGAAAGGCTGGGAAGGGGAGGGAAGGATTCTGGGTGAGAGGCTCATATAGCAG 177814
QY 203 ----- 203
Db 177815 AGCCATGCACTGAGTGAAGGCTGTGTATTATAGCAGACCGGCTGTGCAAGCCTTAG 177874
QY 203 ----- 203
Db 177875 GCCAGCTGGAAGAGGTGAAGCTTGAGAGGGGAGACTGGGGGAGAGGTAGGCTAATGG 177934
QY 203 ----- 203
Db 177935 GGAATGTCCAGGGCAAGTAGCCGTGAGACCCCTTATTATAGCCCACTCAGGGAAACC 177994
QY 203 ----- 203
Db 177995 TCCCAAGTGGCCTTGCAAAAGCAGCTGACGCCAGGAGCCTGGGGAGGTGGAGCAG 178054
QY 204 ----- 204
Db 178055 GCGGCTCAGCTGCTGTCTCTGTGCTTCTTGCGGCGGAGCTCAAGCTGCAAGCGCAAG 178114
QY 210 ProArqIaGluThrIlyPProthrgIuAspLeuArqSerTrpGluYsGlnArqAla 229
Db 178115 CCGGAGCAAGACCAAGCCCACTGAGGACCTGCGGAGCTGGGAGAAAGACGAGCGCC 178174
QY 230 Proleu----- 231
Db 178175 CCGCTCAGTATGTGTGGGACATGTGACATATGTGAGACCAAGCTGTGTACTG 178234
QY 231 ----- 231
Db 178235 CTTCGAAGATTCAAGACAGAAAGTTTCTGTGCTGTGTCTTCAAGGAGCCAA 178294
QY 231 ----- 231
Db 178295 CACTTTGATGACAAGACTCATTAATCAATGAATAATTCAACAAATTAATCATATGTAGT 178354
QY 231 ----- 231
Db 178355 CCAATCCTTTATATGAAACATCTTGGGAGACAGACTGAGGCCAGCATTTTGGG 178414
QY 231 ----- 231
Db 178415 CTTTGAAGGTGCTCTCAGCAGTGCATCAGGCTTGACGCGTTGCTGTCTGTGCGA 178474
QY 231 ----- 231
Db 178475 GATGACTGAATATTACACTAATGATGAGATAAAGCCGTAACGCTTCCATCCCTTAG 178534
QY 231 ----- 231
Db 178535 GGTGGGCTGCTGCTGCATGTGGGTATGAATACTTATGATTTCAGGCTTTTGGACTTC 178594
QY 231 ----- 231
Db 178595 AGAATGTGATTAAGCAATTATAGAGCTGTAGTTACTTTCACCTGTACTTCTTCAAGT 178654
QY 231 ----- 231
Db 178655 GCAAGGAGAGAGGAGGAAAGAAAGAGATTTGTTAAAGAAAGGATTAAGTAACA 178714
QY 231 ----- 231
Db 178715 ACAATACACGCGAGGATTAAGCTGTGTTAACCTTAATGTGTCAATGTGTAAGGAATG 178774
QY 231 ----- 231
Db 178775 ATGGAGCTGGAATTAGTACTGCTAGGGGAAAGAAAAACAAGTTCAAGCAGGGTCAAGC 178834
QY 231 ----- 231

Dh 178835 CAGGTGTGGGACAGGGCTGAGATACAGAGGACCAGGCTGGCTGAGATCTGGGGA 178894
Qy 231 ----- 231
Db 178895 AGGTGTACTGTGTGTGTGGGAAGAACGAGGAGGAGATCGAAGAGTCTTGG 178954
Qy 231 ----- 231
Db 178955 AGGAGAGGCTAGGAGACCCCGAGCCCTGCAGGGGAAAAAGAAAGCCTTGGCAGCATCT 179014
Qy 231 ----- 231
Db 179015 CTCCCAACCCCTCATACACACTGTGTAGTCACATACATGTGCACACACACTCAC 179074
Qy 231 ----- 231
Db 179075 ACATGTGTAGAGCCCATCTGGGCGGTAGGGCAGTGTGGGTCTGTCTGGCCCAACC 179134
Qy 231 ----- 231
Db 179135 CAACAGTTTAAGGGGATTCTGTCCGACACCCCTGAGAACGAGCCAGGGGTTAGGC 179194
Qy 231 ----- 231
Db 179195 CGCTTACTGTGATGGGACATATGTGCTTAAGCTTGGCACTGGGGCATTTCTTGAAGTT 179254
Qy 231 ----- 231
Db 179255 GACCAATTAACCTCTCACGTTTGTGTGAACCTGTTGAGCAAAACCCGTCACCTCTC 179314
Qy 231 ----- 231
Db 179315 CCAGAACCTGCGGGAGGTGGGCTGTGGCTGTGTATGTCTGTCTCTCCATCGGC 179374
Qy 231 ----- 231
Db 179375 AAGGAGGAATGAGGACACAGTTTGTATAGAGACTGTCTCCCAACCCCTCTCC 179434
Qy 231 ----- 231
Db 179435 GAAGGCCAGAGGAATGATGAGGGGTCCAGAGTTCATTCTGGGGGCTTCTTAAGTTGC 179494
Qy 231 ----- 231
Db 179495 CTGTGCGCTTGCCCCACCTGTGTGTGTATCCCACTCACTTGCCTCCCTGCGCT 179554
Qy 232 ArgPhedInglYAspSerArgLeuGlnSerGlyCysTyrHisGlyValAspGlu 251
Db 179555 AGGTTCCAGGGGTACAGCCCGCTGTGACAGTCTGGCTGTACCACTTGGTATGATGAG 179614
Qy 252 AsnIleGluArgArgAsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSerGln 271
Db 179615 AACATGAGAGGAGAAACCACTTATGATCTCGCGGATAGAAAACCTACACGTCTCCAG 179674
Qy 272 PheGlyPro----- 274
Db 179675 TTTGGGCTGTGTAAAGGACTCAAGCACCTTCATGTGGCGATGAGGCGAGGCGTGGCTG 179734
Qy 274 ----- 274
Db 179735 GACGGGCGAGGGGCGGCGCATGACAGTGTGTCTACTGCTTACCCAGGCTTGGTAAACAG 179794
Qy 274 ----- 274
Db 179795 CTATATGGGACAGGTCAAGAACTTTTGGAGAAAGTGAAGTTAAATGGTTTCAAAAT 179854
Qy 274 ----- 274
Db 179855 TAAAGTATATCATGCAGAAATCGAACMAAATCCAAATAGTATCTAAAGGCTTACACCCA 179914
Qy 274 ----- 274
Db 179915 ATGACAAACACAGTTCTGTGCCCCATCCAGGCCCAACCCCAAGCTTCTTCTGGGGGCAAC 179974

Qy 274 ----- 274
Db 179975 CACTTGAACCTCTCTGGCTGTGTCTTCCAGCAAGAACTCGTTCCTTAATGACAGA 180034
Qy 274 ----- 274
Db 180035 CTGCTGTCTTATATCAGCCTTGACATATGTGACCTCTGTTCATTTCACTCCACAA 180094
Qy 274 ----- 274
Db 180095 CTCCCTTCTCCCTTCTCTTCATATGTTAATTCTCTATAGATTACTTGTATGTTT 180154
Qy 274 ----- 274
Db 180155 TTAGCAATATATCTTAGATTCTTCTTCCCGACGTGACCTTAGATATTTTCGGGCTCTC 180214
Qy 274 ----- 274
Db 180215 TGCTGCATGTCTCTTCTCTGGCTCCCTCTCGAGAGTATTTTGAAGACCTTACCAAGA 180274
Qy 274 ----- 274
Db 180275 ACCAGTTCTTCCACTTAGCAGATGGGAAGTTCATTTCAGAGAGTTTCTCTGCAAGT 180334
Qy 274 ----- 274
Db 180335 AGTAGAGGAGATCAGGGCGTGGGGGGTAGCTCTGTGTTTGAGAACTGGGGGCGGGGG 180394
Qy 274 ----- 274
Db 180395 TGATGTGGGGTATAGCGCAAGCCGACAGACACAGTAGTGCTCATTAATGTCTGTTG 180454
Qy 275 -----GlySerProSerValAlaGlnLeuSerGlyLeuProProArgThrSer 290
Db 180455 AATTGTTCTTAGGCTCCCTCCGTGCGCCAGAAAGTCAABAATGCCCCCGGACCTCC 180514
Qy 291 AsnProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgGlyPro 310
Db 180515 AATCCCACTGATCTCGCTCCCATGAGCCGGAAGCCATCCACGTCACACCGGAACCCC 180574
Qy 311 GlnGlyValAspProAlaSerPheHisPheLeuAspProIleAlaIleValSerGlu 330
Db 180575 CAAGGCGTGAACCCGCGCTCTCTTCCACTTCTTGACACCCCAATCGCAAGTCTCAGAG 180634
Qy 331 LeuGlnAlaArgLeuArgGlyThrGlnAspGlySerIleHisPheValArgSerProLys 350
Db 180635 CTCGAGCAAGGCTCCGCGGACCCAGAGACGGAGCAAGCACTTGTGAGGTCCGCCAAG 180694
Qy 351 AlaGlnGlyLysSerValGlyValGlnHisValAlaArgGlyValArgAsnLysProPro 370
Db 180695 GCCCAGGGCAAGAGTGTGGGTGTGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCT 180754
Qy 371 LeuGlyProAlaIleProAlaValSerProSerAlaHisIleuAlaIleAspProAlaLeu 390
Db 180755 CTGGACCGGCAATCCCTGTGGGTGTCCCTTCAACCACTGTGCTGACCCGCGCTTC 180814
Qy 391 LeuProSerLeuAlaProLeuGlnHisIleValLeuHisIleValArgAlaValSerGln 410
Db 180815 CTCCCTCCCTTACCCCTCCGCGGACAGAAACACAGCAACGAGCCAGAGAGGACAG 180874
Qy 411 GlnGlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyIleProValLeuGlyArgGlu 430
Db 180875 CAGGGCTGCCGGGGCTGCAGAGCACACTGGCTTCAAGTGGCCCTGTCTGGGGCGGGAG 180934
Qy 431 HisLeuArgGluLeuProAlaLeuValValTyrGlnSerGlnAlaGlyIleProValGln 450
Db 180935 CACCTGCGGAGCTGCCCGCTTGTGTGTATGAGAGCCAGGCGGGGACGCGGTCCAG 180994
Qy 451 ArgHisGluHisIleHisIleHisIleGluHisIleHisIleTyrHisIleHisPheTyrGlnThr 470
Db 180995 AGACATGAGCACCAACCAACCACTGAACATCAACACCATTAACCAACCTTCTACAGACA 181054

RESULT 13	HSA303140	16319 bp	DNA	linear	PRI 03-APR-2004
LOCUS	HSA303140				
DEFINITION	Homo sapiens NOD2 gene for LRR-containing protein, exons 1-11.				
ACCESSION	AJ303140				
VERSION	AJ303140.1	GI:14488148			
KEYWORDS	LRR-containing protein; NOD2 gene.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1				
	Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tyek, C., O'Morain, C.A., Gasnill, M., Binder, V., Finkel, Y., Corcos, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J.F., Sahbatou, M. and Thomas, G.				
TITLE	Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease				
JOURNAL	Nature 411 (6837), 599-603 (2001)				
MEDLINE	21279172				
PUBMED	11385576				
REFERENCE	2 (bases 1 to 163319)				
AUTHORS	Zouali, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-2001) Zouali, H., Genetics, Fondation Jean Dausset-CEPH, 27 RUE Juliette Dodu, 75010 Paris France, FRANCE				
FEATURES	Location/Qualifiers				
source	1. 163319				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/chromosome="16"				
	/map="16q12-13"				
	/clone_1id="h87b10"				
	/clone_1id="CEPH Human BAC library"				
	98302..131994				
	/gene="NOD2"				
gene	join(98302..98906,106807..106912,109510..111325,11517..115600,115821..115904,118855..118938,121552..121635,122230..122313,124418..124501,128745..128828,130674..131994)				
mRNA	/gene="NOD2"				
	98302..98906				
exon	/gene="NOD2"				
	/number=1				
CDS	join(98448..98906,106807..106912,109510..111325,11517..115600,115821..115904,118855..118938,121552..121635,122230..122313,124418..124501,128745..128828,130674..130746)				
	/gene="NOD2"				
	/function="putative NF-kB activator"				
	/note="protein of the CARD family"				
	/codon_start=1				
	/product="LRR-containing protein"				
	/protein_id="CAC42117.1"				
	/db_xref="GI:14488149"				
	/db_xref="GOA:Q9HC29"				
	/db_xref="Swiss-Prot:Q9HC29"				
	/translation="MCSQEAFOAQRSQLVELLVSGSLGEFESVLDMLSMWEVLSMEDYEGFHLGQPSLHRLRLDLYWNGKWCQGLIAAQBDAQSOPKHLGCMWDHSHLPARDLQSHRPALIVRLSHVENMLDLAMERGFGVGYCDEIRLPIPTPSARARVLLATYVANGLAALFLOHVOELPYPLALPLEAATCKTKYMAKRTTVSAQSRFSTYGAETLCEEDIYENLVAVADVGAGSPQSKPATLGELELSTPQHLNDADTVLVSEAGSGKSTLQRLHMLMAAGDFQEFLLVPFPSCLOCMAPLSVRLTLFELHCCMPVQGE DI FQLLDLHPDPRVLTDFDGFDEFKFRFTDSEBRHCSPTDPSVQTLFNLQGNLKNA RKVTSRPAVASAFIRKIRTFEPNLKGSSEGGIQLYLRKHHEGVADRLRLRLQSL ALHGLCHLPVPSMWSKCHOBLLLOEGSPKTTDMVLLIIOHFLIHAATPRDSASQEL GPRLRGRIPFLHLGRLALWGLCMCCVPSAQDLOAAQVSPDISISGLFVRAGVYPGSTLAPLELHITFQCCFPFAFYLASADVPRLLHLRLNCGRPNSPWARLLPTVICIAS SEGDSSVVALIQAEAPHNLQITAAFLAGLISRHWGLLAEQTSERKALRRQACAM CLASLRKHFFHSIPPAAGGAQSVHAMPGLWLRSLRYEMQEERLARAKAAGLVNCHL KLTFCVGPTEGCALAFVLIQHLRRPVALQLQDYNVSVGIQVYQLLPCIGVCALVLRND				

ORIGIN	NISDRGICKLIECALHCEQLQKLALFNNKLTJDCASHMAKLACRONFLALRLGNYYI TAAGAQVLAEGKVGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSILRMISLVGNISG VGAQALAMIAKRWMLLEPLCLEBNHLODEGVCSLAGIKKNSSLKILKSNNCITTYLG AEALLQHLERNDTLLEVMILRGNTFSLEBVDKLCGRTRLLL"				
	98907..106806				
	/gene="NOD2"				
	/number=1				
exon	106807..106912				
	/gene="NOD2"				
	/number=2				
intron	106913..109509				
	/gene="NOD2"				
	/number=2				
exon	109510..111325				
	/gene="NOD2"				
	/number=3				
intron	111326..115516				
	/gene="NOD2"				
	/number=3				
exon	115517..115600				
	/gene="NOD2"				
	/number=4				
intron	115601..115820				
	/gene="NOD2"				
	/number=4				
exon	115821..115904				
	/gene="NOD2"				
	/number=5				
intron	115905..118854				
	/gene="NOD2"				
	/number=5				
exon	118855..118938				
	/gene="NOD2"				
	/number=6				
intron	118939..121551				
	/gene="NOD2"				
	/number=6				
exon	121552..121635				
	/gene="NOD2"				
	/number=7				
intron	121636..122229				
	/gene="NOD2"				
	/number=7				
exon	122230..122313				
	/gene="NOD2"				
	/number=8				
intron	122314..124417				
	/gene="NOD2"				
	/number=8				
exon	124418..124501				
	/gene="NOD2"				
	/number=9				
intron	124502..128744				
	/gene="NOD2"				
	/number=9				
exon	128745..128828				
	/gene="NOD2"				
	/number=10				
intron	128829..130673				
	/gene="NOD2"				
	/number=10				
exon	130674..131994				
	/gene="NOD2"				
	/number=11				
ALIGNMENT SCORES:					
Pred. No.:	1,06e-46	Length:	163119		
Scores:	1191.50	Matches:	315		
Percent Similarity:	26.36%	Conservative:	1		
Best Local Similarity:	26.27%	Mismatches:	883		
Query Match:	97.68%	Indels:	3		
DB:	9	Gaps:	3		

QY 253 ILEGLUARGHARGHNSHISTYRLEUASPHEALAGLYLLEGLUENTYRTHRSEGLNPH 272
 DB 31117 ATGGAGGAGAAACCACTACTATGATCTGCGGGGATAGAAAACCTACAGTCCCAATTT 31176
 QY 273 GLYPRO----- 274
 DB 31177 GGGGCTGTAAAGGACTCAAGCACCTGCAATGGGCGATGAGGGCAGGGGCTGGAC 31236
 QY 274 ----- 274
 DB 31237 GGGGCGAGCGGCGCGTGGGTGTCTCACTGCTACCCAGGCTTGGTAAGCAACTA 31296
 QY 274 ----- 274
 DB 31297 TTATGGACAGGTCAAGACTCTTGTGAGAAAAGTACGTTAAAATGTTTCAAAATTA 31356
 QY 274 ----- 274
 DB 31357 AGTATATCATGACGAATAATGAAATCAATATGATTAAGGCTTAGACCAATG 31416
 QY 274 ----- 274
 DB 31417 ACAACACAGATTGTTGTGCCCATCCAGCCACCAAGCCGTGTCGCGGCGAACAC 31476
 QY 274 ----- 274
 DB 31477 TTGACCTCTCTGGCTGTTTCTTCCAGCAAGAACTCGGTTCCCTAAGTACAGCACTA 31536
 QY 274 ----- 274
 DB 31537 CTGTGCTTATACAGCTTGACATATGTGACCTTTGCTCAATTGCTACCCCAACTC 31596
 QY 274 ----- 274
 DB 31597 CCTTCTCCCTTCTCTCATATGTTAAATCTCTATGACTTCTGTATGTTTAA 31656
 QY 274 ----- 274
 DB 31657 GCATATATCTTAAGTTCTTTGCCCCCAGTACCTTAGATGTTTGGGGTCTCTGC 31716
 QY 274 ----- 274
 DB 31717 TGCATCTGCTTCTTGGGCTCCCTCGAGAGTATTTTGAAGACTACCAAGAAC 31776
 QY 274 ----- 274
 DB 31777 AGTTCTTCCACTTAGACAGATGGGAAATTCATTTCAGAGATTTCTGCAAGCAGT 31836
 QY 274 ----- 274
 DB 31837 GAGGGGACGTCAAGGCACTGGGGGGTAGCTGTGTTTGAAGAACTGGGGGGGGGTGA 31896
 QY 274 ----- 274
 DB 31897 TGTCTGGGTATAGCGCAAGCCGAGACACAGTAGGAGTCAATTAATGCTGTGAT 31956
 QY 275 ----- 275
 DB 31957 TGGTCTTAGGCTCCCTTCGTGCGCCAGAAAGTCAAAACTGCCCCCGGCACTTCAAT 32016
 QY 292 PROTHARGSERARGSERHISGLUPROGLUALALHISILEPROHISARGLYSARGGLN 311
 DB 32017 CCCACTGATCTCGCTCCCATAGCCGAAAGCCATCCACATCCACACCGAAAGCCCCCA 32076
 QY 312 GLYVALASPProalaserPhehisPheLeuaspThrProilealysValserGluLeu 331
 DB 32077 GGGGTGAGCCCGGCTCTTCACTTCTTACACCCCAATCGCCAAAGTCTCAGAGCTC 32136
 QY 332 GLINGLARGLEUARGLYTHRGINASPGLYSEIRYSHISPHETVALARGSERPROLYSALA 351
 DB 32137 CAGCAAGCGGCTCGGGGACCCCAAGAGCGGAGCAAGCACTTGTGAGGTCCCCCAAGGCC 32196

QY 352 GLINGLYSERVALGLYVALGLYHISVALALAAARGGLYALAARGHANSUYPROLEU 371
 DB 32197 CAGGGCAAGGTGTGGGTGTGGGCCACGTGGCCAGAGGGGCAAGAACGCCCTCTG 32256
 QY 372 GLYPROALALAEProalavalSerProserAlahisLeuAlaAserProalaleuLeu 391
 DB 32257 GAGACCGGCATCCCTCGGTGTCCCTCCGCGCCACTGCTCCAGCCCGGCTCTC 32316
 QY 392 PROSERLEUALPROLEUGLYHISLYSLYSHISLYSHISARGVALALYSGLUSERGLN 411
 DB 32317 CCTCTCTAGCCCCCTCGGGCAAGAACACAGACCGAGCCAGAGAGCCAGCAG 32376
 QY 412 GLYCYARGGLYLEUGLINALPROLEUALASERGLYGLYPROVALLEUGLYARGLYHIS 431
 DB 32377 GGTGCGGGGCTGTGAGGACACACCTGCTCAGTGCGCTGTCTCGGGGCGGAGCAC 32436
 QY 432 LEUARGLYLEUProalaleuValValTYRGLISERGLNALAGLYINPROVALGINARG 451
 DB 32437 CTGCGGAGCTGCGCGCTTGTGTGTATGAGAGCCAGGCCGCGGACGCGTCCAGAGA 32496
 QY 452 HISGLUHHISHISHISHIGHISHISHISHISHISHISHISHISHISHISHISHISH 470
 DB 32497 CATGAGCACCAACCACTGAACATCACCATTTACCACTTACCAAGACA 32553
 RESULT 14
 AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
 LOCUS
 DEFINITION Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
 AC007608
 VERSION AC007608.7 GI:29124038
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 169773)
 Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesner,J., Meincke,L., Lomnixe,J., White,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bused,M., Sutherland,R., McMurry,K., Han,C. and Deaven,L.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M688, Los Alamos, NM 87545, USA
 3 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 5 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Mar 19, 2003 this sequence version replaced gi:18129388.
 COMMENT Draft Sequence Produced by DOE Joint Genome Institute
 www.igi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
 www-shgc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;

FEATURES Estimated Total Number of Errors is 0.
Location/Qualifiers

1.169773
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-401P9"

ORIGIN

Alignment Scores:

Score: 1.11e-46 Length: 169773
Pred. No.: 1191.50 Matches: 315
Percent Similarity: 26.36% Conservative: 1
Best Local Similarity: 26.27% Mismatches: 1
Query Match: 47.68% Indels: 883
DB: 9 Gaps: 3

US-09-993-966-7 (1-470) x AC007608 (1-169773)

QY 154 GluAp1leThrSerLeuLeuHisThrIleTygIuValaAapSerSerValaAsnHis 173
Db 81532 CAGGACATCAACAGCTTGTGCAACCATCTATAGAGTGTGACCTCTGTCAACAC 81473
QY 174 SerProThrSerSerLeuSerMetLeuArgValIleLeuThrValaIaProaAspGlySerGln 193
Db 81472 TCCCAACATCCAGCAAGATGCTGGGGTAAAGCTCACTGCGCCCGATGGACGAC 81413
QY 194 SerLeuArgSerValLeuValaAsnGlnaA----- 203
Db 81412 AGCAAGAGGAGCGCTCTGTCAATCAAGCTGTGAGAGGCTGCAAGGCTGAGGCTGGAAA 81353
QY 203 ----- 203
Db 81352 CAATGTCCTCCATCTCAGAAAGAACAGACCTTCTGGAGGCGGAAAGGAACTGTG 81293
QY 203 ----- 203
Db 81292 TGGGCTTTCTGGGAGAGCTGTTGTGACAGGTGGGTTCAATGAGTCTTGAAAGG 81233
QY 203 ----- 203
Db 81232 TAGAGAGAGGTGGGAAGGAGGAGGAAAGGATTCTGGTGAGAGGCTCATATAGGCA 81173
QY 203 ----- 203
Db 81172 AGCCATGACAGTGAATGAAGAGTGTGTTACGGGACAGATGGCTGTGCAAGCTTGAG 81113
QY 203 ----- 203
Db 81112 GCCAGCTGAAGAGAGTGAAGCTGGAGGGAGTCTGGGGAGAGAGGTAGAGTGAATGG 81053
QY 203 ----- 203
Db 81052 GGATGTGTCCAGGGCAGGTAGCTTGAGACCCCTTTATTAGCCCCAGCTCAGGGAACCC 80993
QY 203 ----- 203
Db 80992 TCCACAGTGCGCTTGCAAAAGCAGCTCAGCGCAAGCAAGCCCTGGGAGGGTGGAGACAG 80933
QY 204 ----- 204
Db 80932 GCAGCTCAGCTGTGTCTCTGTGCTTCTGGGCGGAGCTCAGACCTGCGAGAGGCAAG 80873
QY 210 ProArgIaIaIuThrIuSProThrGluAapLeuArgSerTrpGluIuSbIuSgIaArgA 229
Db 80872 CCGGAGCAAGACCAAGCCCACTGAGGACCTGGGAGCTGGAGAAAGAAAGCAGCGAGCC 80813
QY 230 ProLeu----- 231
Db 80812 CCGCT-CAGGTATGTGGCATGTGACATGACATATGTTAGACCAAGCTGTGACT 80754
QY 231 ----- 231

Db 80753 GCTTCTGAAGGATTCAAGCAGGAAGTTTGTCTGTGTCTGTCTTCTTCAGGAGCCA 80694
QY 231 ----- 231
Db 80693 AACTTTTGAATGCAAGACTCATTAATCAATGAATAATCTTAACAATAACTACATGTAG 80634
QY 231 ----- 231
Db 80633 TCCACATCTTTTATGGAACATCTTGGGACAGAGACTGAGGTACAGATTTTTGG 80574
QY 231 ----- 231
Db 80573 GCTTGAAGAGTGTCTCTCAGAGTGCATAGGCTTGAAGTGTGTCTTCTGTGG 80514
QY 231 ----- 231
Db 80513 AGATGACTGAATATTCACTAAGTGAATGAACCGTAACACCTTCCATCCCTTTA 80454
QY 231 ----- 231
Db 80453 GGGTGGGCTGTGCTGCCAGCTGGGTATGAATACTTATGTTTCAAGCTTTTGAAGCT 80394
QY 231 ----- 231
Db 80393 CAGATGTGAATGAAGAAATTATGAGCTGTATACCTTTCACTGTATCTTCTTCAAC 80334
QY 231 ----- 231
Db 80333 TGCAGGAGGAAGGAAGGAAGAAAGAGATTGTTAAAGAAAGGATTGAGTAAC 80274
QY 231 ----- 231
Db 80273 AAGATTAACAGCGAGCATTAAGCTGTGTAACTCTAGTGTCAATGTAAGGAAT 80214
QY 231 ----- 231
Db 80213 GATCGAGCTGATTAGTACTCGCTAGGGGAAAAAAGAAACAGTTCAAGAGGTCAAG 80154
QY 231 ----- 231
Db 80153 CCAAGTGTGGGACAGGCTGAGAAATCAGAGGACCCAGGCTGCTAGAGATCCTGGGG 80094
QY 231 ----- 231
Db 80093 AAGTGTACTGTGTGTGTGTGGAAGGAAGCAGAGGAGCAGATGGAAGGTCTTG 80034
QY 231 ----- 231
Db 80033 GAGGAGAGGCTAGGAGGCCGAGCCCTGAGGGGAAAAAAGAAAGCTTGGCAGCATC 79974
QY 231 ----- 231
Db 79973 TCTCCCAACCCCTCATGACACTGTGTAGTACACATGATGTGACACACACTCA 79914
QY 231 ----- 231
Db 79913 CACACTGTGAGGCCATCTGGGCGTGAAGGAGGTATGAGGTCTGTGCGCAAC 79854
QY 231 ----- 231
Db 79853 CCAAGATTTAAGGGGATTTGTGTGCCACACCCCTGAGAAACAGGCCAGGGTTAGG 79794
QY 231 ----- 231
Db 79793 CCGCTTACTGACGTGGGACATATAGCTTGAAGCTTGGACCTGGGCAATTCCTAGAAGT 79734
QY 231 ----- 231
Db 79733 TGAACAATACTCTCAGCTTGTCTTTGAACCTGTTCAGAGCAAAACCCGTCACCTCT 79674
QY 231 ----- 231

Db 79673 CCCAAGAGCTGCGGGAGAGTCTGGGCTGTGGGTATAGTCTGTCTTCATCGG 79614
 QY 231 ----- 231
 Db 79613 CAAGAGGAAATGAGGACAGTTCTGTATAGAGACTGTCTCCCACTCTCCAA 79554
 QY 231 ----- 231
 Db 79553 GGCAGAGAAATGAAATGAAGGCTCAGAGTTCACTTCGGGGCTTCTTAGAGCTA 79494
 QY 232 -----Arg 232
 Db 79493 TGGCTTGCCTCCCACTGTGTGTGTATCCCACTCACTTCCTCCCTGCGTAGG 79434
 QY 233 PheGlnGlyAspSerAlaGluGlnSerGlyCysArgThrHisCysValAspGluAsn 252
 Db 79433 TTCGAGGTATACAGCCCTCGAGACAGTCTGGCTGTACCACTTCGAGATGAAAC 79374
 QY 253 IleGluArgArgAsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSerGlnPhe 272
 Db 79373 ATCGAGGAGAAACCACTTAGATCTGCGCGGATAGAAACTACAGTCCCAATTT 79314
 QY 273 GlyPro----- 274
 Db 79313 GGGCTGTGAAGGACTCAAGCACCTGCAATGGCGATGAGGCGAGCGTGGCTGAC 79254
 QY 274 ----- 274
 Db 79253 GGGCGAGCGGGCCGTGGGTGTGTCTGCTACCCAGCTTCGGTAAACAATA 79194
 QY 274 ----- 274
 Db 79193 TTATGGACAGGTCAAAAGACTTCTTGAGAAAGTACGTTAAATGTTCAAAATTA 79134
 QY 274 ----- 274
 Db 79133 AGTAAATCATGACGAATTCGAACAAATCCATATGATCTAAAGGTTAGACCAATG 79074
 QY 274 ----- 274
 Db 79073 ACAAAACAAGTTTGTGCCCATCCAGCCCAAGCCGTGTCGGGGGCAACAC 79014
 QY 274 ----- 274
 Db 79013 TTTGACCTCTCCGGCTGTTTCTTCCAGCAAGAACTCGGTTCCCTAAGTACAGACTA 78954
 QY 274 ----- 274
 Db 78953 CTGTGTCTTATCAGCCTTGACATATGTGACCTCTTGCAATTTGACTCCCAACTC 78894
 QY 274 ----- 274
 Db 78893 CTTTCTCCCTTCTCTCATATGTTAAATCTCTATGAGTACTTCTGTATGTTTTTA 78834
 QY 274 ----- 274
 Db 78833 GCAATATACTTAGAGTCTTCTTGTGCCCAAGTACCTTAGATGTTTCCGGGCTCTGC 78774
 QY 274 ----- 274
 Db 78773 TGCATCTGTCTTCTGCGCTCCTCTCGAGAGTATTTTGAAGACTACCAAGAAC 78714
 QY 274 ----- 274
 Db 78713 AGTCTTCCCACTTAGAGATGGAAAGTTCCATTTAGAGAGTTTCTGCAAGGAGT 78654
 QY 274 ----- 274
 Db 78653 GAGGGGAGTCAGGGCATTTGGGGGTAGCTTGTTTTGAGAACTGGGGGCGGGGTGA 78594
 QY 274 ----- 274
 Db 78593 TGTCTGGGTATAGCGCAAGCCCAAGACACAGTAGGTGTCATTAATGTCTGTGAAT 78534

QY 275 -----GlySerProSerValAlaGlnIleValSerGluLeuProProArgThrSerAsn 291
 Db 78533 TGGTCTTAAAGCTCCCTTCTCGTGGCCAGAAAGTCAAGATTGCCCCCGACCTCCAA 78474
 QY 292 ProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysProGln 311
 Db 78473 CCACCTGATCTGCTGCCATCGAGCCGGAAGCATCCATCCACACGAAAGCCCCAA 78414
 QY 312 GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeu 331
 Db 78413 GCGGTGACCCCGGCTCTTCCACTTCCTTGACACCCCAATCCCAAGTCTCAGACTC 78354
 QY 332 GlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisIlePheValArgSerProLysAla 351
 Db 78353 CAGCAAGCTCTCGGGGACCCAGAGCGGAGCAAGACATTTGTGAGTCCCCCAAGGCC 78294
 QY 352 GlnGlyLysSerValGlyValGlyHisValAlaArgGlyValaArgAsnLysProProLeu 371
 Db 78293 CAGGGCAAGAGTGTGGGTGTGGGCCAAGTGGCCAGAGGGGCAAGAAACAAGCCCCCTG 78234
 QY 372 GlyProAlaIleProAlaValSerProSerAlaHisLeuAlaAspProAlaLeu 391
 Db 78233 GAACCGCCATCCCTGCGGTGTCCTCCCGCCACCTGCTCCAGCCCGGCTCTC 78174
 QY 392 ProSerLeuAlaProLeuGlyHisIleLysLysHisIleLysHisIleArgAlaLysGluSerGln 411
 Db 78173 CCTTCTTACCCCCCTCGGGCAAGAGACACAGACACCGAGCCAGAGACCCAGCAG 78114
 QY 412 GlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgLysHis 431
 Db 78113 GGTGCGGGGCGCTGCGAGGACACATGACATGACCATTCACCATTTACCAAGCA 78054
 QY 432 LeuArgGluLeuProAlaLeuValValTyrGluSerGlnAlaGlyGlnProValGlnArg 451
 Db 78053 CTGCGGAGCTGCGCCGCTTGTTGTGTATGAGAGCGAGCCGCGGCGGCTCAAGA 77994
 QY 452 HisGluHisHisHisHisHisGluHisHisHisHisIleTyrHisHisIlePheTyrGlnThr 470
 Db 77993 CATGAGCACCAACACACATGACATGACCATTCACCATTTACCAAGCA 77937

RESULT 15
 LOCUS AC007334/c 170791 bp DNA linear PRI 13-MAR-2003
 DEFINITION Homo sapiens chromosome 16 clone R11-147B17, complete sequence.
 ACCESSION AC007334
 VERSION AC007334.8 GI:28933524
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 170791)
 DOE Joint Genome Institute, Stanford Human Genome Center and Los
 Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 170791)
 Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
 Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
 Campbell,C., Fawcett,J., Maltbie,M., Bussed,M., Sutherland,R.,
 McMurry,K., Han,C. and Deaven,L.
 DIRECT SUBMISSION
 Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 3 (bases 1 to 170791)
 DOE Joint Genome Institute.
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 JOURNAL 4 (bases 1 to 170791)
 REFERENCE


```

Db      29398 CCAACAGTTTAAGGGATTCGTGTCGCCACACCCCTGAGGAACGAGGCCAGGGGTTAGG 29339
QY      231 ----- 231
Db      29338 CCGCTACCTGCAGTGGGACATATGAGCTTGAAAGCTGGGACCTGGGACATTCTTAGAGT 29279
QY      231 ----- 231
Db      29278 TGAACATTAACCTCTCAAGTTTGTCTTGAACCTGTTCAGAGCAAAACCCGTCACACTCT 29219
QY      231 ----- 231
Db      29218 CCCAAGAGCTGCGGGAGAGTTCGGGCTGTGGCTATGAGTCTGTCTCTCATCGG 29159
QY      231 ----- 231
Db      29158 CAAGAGGGAATGAGAGCACAGTTGATATAGAGAGACTGTTCTCCACCTCTCCAAA 29099
QY      231 ----- 231
Db      29098 GGGCAAGAGAATAATGAATGAAAGGTCACAGATTCTTGGGGGCTTCCTAGTAGCTA 29039
QY      232 ----- Arg 232
Db      29038 TGGCCTTGCCCCCAGCTGTGTGTGTGTATCCCACTCACTTCTCCCTGCGGTAGG 28979
QY      233 PheGlnGlyAspSerAlaGluGlnSerGlyCysValAspGluAsn 252
Db      28978 TTCCAGGGGTACACGCCCTGAGACAGTGTGGCTGCTACCAACATTCGATAGTAAGAAC 28919
QY      253 ILeuArgArgAsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSerGlnPhe 272
Db      28918 ATCGAGAGGAAACCACTACTTAGATCTCGCCGGATAGAAAACCTACAGCTCCCAATTT 28859
QY      273 GlyPro----- 274
Db      28858 GGGCTGTGAAGGAGCTCAAGACCTGCAATGGGCGATGAGGAGGCGTGTGCTGAC 28799
QY      274 ----- 274
Db      28798 GGGCGAGGCGGGCCGTGGGGGTGTCTACTGCTACCCAGGCTTGGTAGAACAACTA 28739
QY      274 ----- 274
Db      28738 TTATGGGACAGGTCAAAAGACTTCTTGAGAAAGTAGCTTAAATGGTTTCAAAATTAA 28679
QY      274 ----- 274
Db      28678 AGTATACATGACGAATCGAACAAATCCATATGATCTAAGAGCTTAGACCCCAATG 28619
QY      274 ----- 274
Db      28618 ACAAACAAGTTTGTGCCCCCATCCAGCCCAAGCCGTGTCCCGGGGCAACAC 28559
QY      274 ----- 274
Db      28558 TTTGACCTCTCGGCTGTTTCTTCCAGCAAGAACTCGGTTTCCCTAAGTACAGACTA 28499
QY      274 ----- 274
Db      28498 CTGTGTCTTATCAGCTTGACATATGTTGACTCTTGCTCAATTTGACCCACAATC 28439
QY      274 ----- 274
Db      28438 CTTTCTCCCTTCTCTTCAATATGTTAATCTCTATGAGTTACTTCTGTATGTTTTTA 28379
QY      274 ----- 274
Db      28378 GCAATATACTTAGAGTCTTTCTTGTGCCCAAGTACCTTAATAGTTTTCGGGGTCTCTGC 28319
QY      274 ----- 274
Db      28318 TGGCATCTGTCTTCTGTGGGCTCCCTCTCGAGAGTATTTTGAAGACCTACAAAGAAC 28259

```

```

QY      274 ----- 274
Db      28258 AGTTCTCCCACTTAGCAGATGGGAAAGTTCCATTTCAAGAGTTTTCGCAAGGAGT 28199
QY      274 ----- 274
Db      28198 GAGGGGAGTCAGGGCATTTGGGGGTAGCTGTGTTTGGAGAACTGGGGCGGGGTGA 28139
QY      274 ----- 274
Db      28138 TGTCTGGGATATAGCGCAAGCCCAAGCACACAGTAGGTGCTCATTAATGTCGTGAAT 28079
QY      275 -----GlySerProSerValAlaGlnLysSerGlyLeuProProArgThrSerAsn 291
Db      28078 TGGTCTTAGGCTCTCCCTTCGTCGGCCAGAAATCGCAATGCCCCCGCACCTCCAAT 28019
QY      292 ProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysProGln 311
Db      28018 CCCACTGATCTCGCTCCCATGAGCCGGAAGCCATCCATCCACACCGAAAGCCCAA 27959
QY      312 GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeu 331
Db      27958 GGGGTGACCCGGCTCTCTTCACTTCTTGAGACCCCAATCCCAAGTCTTAGAGCTC 27899
QY      332 GlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysAla 351
Db      27898 CAGCAACGGCTCCGGGGCACCCAGACGAGAGCAAGACACTTTGTAGAGTCCCCCAAGCC 27839
QY      352 GlnGlyLysSerValGlyValGlyHisValAlaArgGlyValaArgAsnLysProLeu 371
Db      27838 CAGGGCAAGAGTGTGGGTGTGGGCCAAGTGGCCAGAGGGGCAAGAAACAAGCCCCCTGT 27779
QY      372 GlyProAlaIleProAlaValSerProSerAlaHisIleuAlaIleSerProAlaLeuLeu 391
Db      27778 GAACCGGCATCCCTGGGGGTGTCCTCCGCGCACTGGCTGCAAGCCGGGCTCTC 27719
QY      392 ProSerLeuAlaProLeuGlyHisLysLysHisLysHisIleArgAlaLysGluSerGln 411
Db      27718 CCTCTCCTAGCCCCCTCGGGGCAAGAGACACAGACCGAGCCAGAGAGCGACAG 27659
QY      412 GlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGluHis 431
Db      27658 GGCCTGCGGGGCTGACAGCACCACTGGCTCAGTGGCCCTGTCTGGGGCGGAGACAC 27599
QY      432 LeuArgGlyLeuProAlaLeuValValTyrGlnSerGlnAlaGlyGlnProValGlnArg 451
Db      27598 CTGGGGAGCTGCCCGCTTGGTGTATGAGAGCCAGGCGGGGACGCGTCCAGAGA 27539
QY      452 HisGluHisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db      27538 CATGAGCACACACCAACCATGAATCATCACCATTTACACACTTTTACAGACA 27482

```

Search completed: December 30, 2004, 06:25:17
 Job time : 8483 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 09:35:26 ; Search time 931 Seconds
(without alignments)

2851.429 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 470
Sequence: 1 MGKLSKPAVCKRRESPEC.....RHEHHHHHHHHHHFYQT 470

Scoring table:

Gapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 4176236 seqs, 2824127955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8342757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-Q/cgnt2_1/USPTO.spool.p/US0993966/runat.29122004.130830.9445/app.query.fasta_1.647
-DB=Published Applications NA -QFMT=fasta -SUFFIX=olg.rnpb -MINMATCH=0.1
-LOPFCU=0 -LOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.csl -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=tbl -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0993966 @CGN 1.1 912 @runat.29122004.130830.9445
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAP=60 -XGAPEXT=60
-Fgapop=6 -Fgapext=7 -Ygapop=60 -Ygapext=60 -Delop=6 -Delext=7

Database: Published Applications NA:*

1:	/cgnt2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgnt2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
3:	/cgnt2_6/prodata/1/pubpna/US05_PUBCOMB.seq:*
4:	/cgnt2_6/prodata/1/pubpna/US04_PUBCOMB.seq:*
5:	/cgnt2_6/prodata/1/pubpna/US03_PUBCOMB.seq:*
6:	/cgnt2_6/prodata/1/pubpna/US02_PUBCOMB.seq:*
7:	/cgnt2_6/prodata/1/pubpna/US01_PUBCOMB.seq:*
8:	/cgnt2_6/prodata/1/pubpna/US00_PUBCOMB.seq:*
9:	/cgnt2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10:	/cgnt2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
11:	/cgnt2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
12:	/cgnt2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
13:	/cgnt2_6/prodata/1/pubpna/US05_PUBCOMB.seq:*
14:	/cgnt2_6/prodata/1/pubpna/US04_PUBCOMB.seq:*
15:	/cgnt2_6/prodata/1/pubpna/US03_PUBCOMB.seq:*
16:	/cgnt2_6/prodata/1/pubpna/US02_PUBCOMB.seq:*
17:	/cgnt2_6/prodata/1/pubpna/US01_PUBCOMB.seq:*
18:	/cgnt2_6/prodata/1/pubpna/US00_PUBCOMB.seq:*
19:	/cgnt2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
20:	/cgnt2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
21:	/cgnt2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	1448	13 US-10-087-192-179	Sequence 179, App
2	470	100.0	1786	10 US-09-993-966-1	Sequence 1, Appl
3	470	100.0	1859	10 US-09-993-966-5	Sequence 5, Appl
4	196	41.7	2379	10 US-09-764-891-5802	Sequence 5902, Ap
5	196	41.7	2379	10 US-09-764-891-5504	Sequence 5904, Ap
6	196	41.7	106315	13 US-10-087-192-178	Sequence 178, App
7	195	41.5	590	10 US-09-993-966-21	Sequence 21, Appl
8	138	29.4	427	10 US-09-918-995-8686	Sequence 8686, Ap
9	49	10.4	148	10 US-09-993-966-18	Sequence 18, Appl
10	49	10.4	1401	9 US-09-730-989-1	Sequence 1, Appl
11	49	10.4	1401	10 US-09-993-966-6	Sequence 6, Appl
12	49	10.4	1416	10 US-09-993-966-2	Sequence 2, Appl
13	49	10.4	2807	13 US-10-087-192-176	Sequence 176, Appl
14	46	9.8	1950	13 US-10-016-1574-126	Sequence 126, App
15	45	9.6	598	15 US-10-029-386-7990	Sequence 7990, Ap
16	44	9.4	133	15 US-10-029-386-21690	Sequence 21690, A
17	44	9.4	134	10 US-09-993-966-14	Sequence 14, Appl
18	42	8.9	128	10 US-09-993-966-20	Sequence 20, Appl
19	36	7.7	639	15 US-10-243-552-231	Sequence 231, App
20	35	7.4	107	10 US-09-993-966-16	Sequence 16, Appl
21	32	6.8	96	10 US-09-993-966-17	Sequence 17, Appl
22	32	6.8	98474	13 US-10-087-192-175	Sequence 175, App
23	29	6.2	554	15 US-10-029-386-7615	Sequence 7615, Ap
24	28	6.0	179	15 US-10-029-386-21315	Sequence 21315, A
25	27	5.7	85	10 US-09-993-966-19	Sequence 19, Appl
26	22	4.7	67	10 US-09-993-966-15	Sequence 15, Appl
27	22	4.7	566	13 US-10-027-633-266931	Sequence 266931,
28	22	4.7	566	15 US-10-027-633-266931	Sequence 266931,
29	22	4.7	2724	9 US-09-833-981-1214	Sequence 1214, Ap
30	22	4.7	2724	9 US-09-833-981-1215	Sequence 1215, Ap
31	17	3.6	1817	9 US-09-822-849A-83	Sequence 83, Appl
32	17	3.6	1959	18 US-10-723-860-5220	Sequence 5220, Ap
33	17	3.6	1966	9 US-09-764-868-152	Sequence 152, App
34	17	3.6	73088	17 US-10-322-696-37	Sequence 37, Appl
35	15	3.2	674	13 US-10-027-633-12063	Sequence 12063, A
36	15	3.2	674	15 US-10-027-633-12063	Sequence 12063, A
37	14	3.0	901	13 US-10-027-633-162722	Sequence 162722,
38	14	3.0	901	15 US-10-027-633-162722	Sequence 162722,
39	13	2.8	488	9 US-09-783-590-676	Sequence 676, App
40	11	2.3	34	10 US-09-993-966-13	Sequence 13, Appl
41	10	2.1	574	9 US-09-864-761-228	Sequence 228, App
42	10	2.1	669	9 US-09-864-761-17051	Sequence 17051, A
43	9	1.9	27	9 US-09-730-989-17	Sequence 17, Appl
44	9	1.9	296	15 US-10-317-444-241	Sequence 241, App
45	9	1.9	236	15 US-10-317-444-242	Sequence 242, App

ALIGNMENTS

RESULT 1
US-10-087-192-179
; Sequence 179, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 179
; LENGTH: 1448
; TYPE: DNA

ORGANISM: Homo sapiens
US-10-087-192-179

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	470.00	470	0	0	0	0
Percent Similarity:	100.00%					
Best Local Similarity:	100.00%					
Query Match:	13					

US-09-993-966-7 (1-470) x US-10-087-192-179 (1-1448)

```

QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLeuArgArgGlySerProGly 20
Db 39 ATGGGGAAACTTCACTCAAGCCGGCCGCTGTGCAAGCGAGAGCCGGAAAGGT 98
QY 21 AspSerPheAlaValSerAlaAlaTrpAlaArgLysGlyIleGlyIleGlyIleGly 40
Db 99 GACAGCTTCCCGTGAACGCTGCTGGGCTCGAAGGCGATCGAGAGTGGATGGAGA 158
QY 41 GlnArgCysProGlyIleValSerGlyProArgGlyLeuArgLeuAlaGlyThrIleGly 60
Db 159 CACGCGTCCCGGCGGCTGTCTCGGAGACCCGACAGCTGCGGTTGCGGCGACCATAGGC 218
QY 61 ArgSerThrArgGlyLeuValAlaValAspValLeuArgSerThrLeuSerGlyIleGly 80
Db 219 CGAAGCACCCGGAGAGCTGTGGCGAGCTGTGAGACACAGCTCAACGAGAAAGAGAG 278
QY 81 AspAspPheArgLeuGlyValAlaLeuProProGlyIleThrAspGlyLeuGlySerGly 100
Db 279 GAGGACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAAAGCTGACGGGCTGGGCGAG 338
QY 101 AspGlyLysLysMetGlyValGlyValSerGlyProCysProGlySerLysLysGlyLeu 120
Db 339 GATGAGAAAGAGTGAAGAGTGAAGAAACCTCGCCAGGCTTCAAGAAAGCAGCTGAG 398
QY 121 PheGlyIleLeuGlyIleCysAspValSerMetGlyIleLeuAspSerArgGlyIleThrPhe 140
Db 399 TTGGAAGGCTCCAGTCCAGTGTCCATGAGAGAGACAGCCGGAGAGTGAACCTTC 458
QY 141 ThrLeuThrAspPheAspAsnGlyLysValThrArgLysPheIleThrSerLeu 160
Db 459 ACCGTGATGACTTGTGACAAACAGGAGGTCAACCGAGAGGACATCACAGCTTCTG 518
QY 161 HisThrIleArgGlyValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 519 CACACCATCTATGAGTGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAGATG 578
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlyIleSerLysArgSerValLeu 200
Db 579 CTGGGGTAAAGCTCACCGTGGCCCCCGATGGAGCCAGCAAGCAAGAGAGGTCTTGTG 638
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlyIleThrLysProThrGlyAspLeu 220
Db 639 AATCAGGCTGACTCGCAGAGGCGAAGGCCCGAGCAGACAGCAAGCCCATGAGGAGCTG 698
QY 221 ArgSerTrpGlyLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGly 240
Db 699 CGGAGCTGGGAGAAAGAGCAGCAGCCCGCTCAGTTCCAGGCTGACAGCCGCTGAG 758
QY 241 GlnSerGlyCysThrHisCysValAspGlyLysGlnIleGlyIleGlyIleGlyIleGly 260
Db 759 CAGCTTGGCTGTACCAACCATGCTGTAGATGAGAAACATCGAGAGAAACCACTACTTA 818
QY 261 AspLeuAlaGlyIleGlyIleGlyIleThrSerGlnPheGlyProGlySerProSerValAla 280
Db 819 GATTCGCCCGGAGTAAAGAAATACATCACTCCCAATTGGGCTGCTCCCTTCCTGGCC 878
QY 281 GlnLysSerGlyLeuProProArgThrSerAsnProThrArgSerArgSerHisGlyPro 300
Db 879 CAGAAAGTCAAGAACTGCCCGCCAGCTCCATCCATCCGATTCCTCCATGAGCCG 938

```

```

QY 301 GluAlaIleHisIleProHisArgLysProGlyIleValAspProAlaSerPheHisPhe 320
Db 939 GAGCCATCCATCTCCACACCGAAGCCCGAAGGCTGGAGCCGCTCTTCCACTTC 998
QY 321 LeuAspThrProIleAlaLysValSerGlyLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
Db 999 CTTGACACCCCATTCGCCAAGGTCTCAGAGCTCACAAACGCTCCGGGCGACCCAGAGAC 1058
QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHis 360
Db 1059 GGGAGAGACACTTGTGAGTCCCGCAAGGCCCAAGGAGAGTGGGTGGGCGCAC 1118
QY 361 ValAlaArgGlyValAlaArgLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1119 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGTGGACCCGCTGCGGTGTCCCCC 1178
QY 381 SerAlaHisLeuAlaAspProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1179 TCCGCCACCTGGCTGCAGCCCGGCTCTCCCTCCCTAAGCCCGGCGACAAAG 1238
QY 401 LysHisLysHisArgAlaLysGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1239 AAGCAGAGCAGCAGAGCCAGAGAGCCAGAGAGCTGCCGGGCTTGACAGCAACCTG 1298
QY 421 AlaSerGlyLysProValLeuGlyValArgGlyHisLeuArgGlyLeuProAlaLeuVal 440
Db 1299 GCCTCAGTGGCCCTGTCTGGGCGGAGCACTCGGAGCTGCCCGCTGTGTGTG 1358
QY 441 TyrGlySerGlnAlaGlyIleProValGlnArgHisGlyHisHisHisHisGlyHis 460
Db 1359 TATGAGAGCAGGCGCGGCGAGCGGTCCAGAGACATGAGACCAACCAACATGAACAT 1418
QY 461 HisHisHisThrHisHisPheThrGlnThr 470
Db 1419 CACCACCATTAACCACTTCAACAGACA (448)

```

RESULT 2

US-09-993-966-1

Sequence 1, Application US/09993966
Publication No. US20030186232A1

GENERAL INFORMATION:

APPLICANT: ROHAN, MICHAEL

TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF

FILE REFERENCE: 014024/0280733

CURRENT APPLICATION NUMBER: US/09/993, 966

CURRENT FILING DATE: 2001-11-27

PRIOR APPLICATION NUMBER: 60/252, 884

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/291, 109

PRIOR FILING DATE: 2001-05-16

PRIOR APPLICATION NUMBER: 60/325, 571

PRIOR FILING DATE: 2001-10-01

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 1786

TYPE: DNA

ORGANISM: Homo sapiens

US-09-993-966-1

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0	1786	470	0	0	0
Percent Similarity:	470.00					
Best Local Similarity:	100.00%					
Query Match:	100.00%					

US-09-993-966-7 (1-470) x US-09-993-966-1 (1-1786)

```

QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLeuArgArgGlySerProGly 20

```


Db 464 GAGACCTTGGCGTGAAGTGGCCCTCTCTGAGAAAGCTGACGGGCTGGGACGGGA 523
Qy 101 AspgulublybWetGluValSerGluProCySPROGlySerlyblyGlnLeuLys 120
Db 524 GATGAGAAAGAGTGAAGAGTGAAGAACCTTGCCAGGCTCCAGAAAGCACTGAG 583
Qy 121 PheGluGluLeuGlnCySPROGlySerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db 584 TTGGAAGAGCTCCAGTGGAGCTGCTCAATGAGAGGACGCGGACGAGTGAACCTTC 643
Qy 141 ThrLeuTyraSPheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeu 160
Db 644 ACCCTGATGACTTTGACAAACAGGCAAGGTCAACCGAGAGGACATCACAGCTTGCTG 703
Qy 161 HisThrIleTyrgluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 704 CACACCATATATGAGTGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATG 763
Qy 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
Db 764 CTGGGGGTAAGCTCACCGTGGCCCGATGGCAGCCAGACAAAGAGAGGCTCTGTGTC 823
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
Db 824 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGAGACAGACCAAGCCCATCAGAGACTG 883
Qy 221 ArgSerTrpGluLysblyGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 884 CGGAGCTGGGAGAAAGAGCGAGCGCCCGCTCAGGTTCCAGGCTGACAGCGCTGAG 943
Qy 241 GlnSerGlyCySPROGlySerHisCysValAspGluAsnIleGluArgAspAsnHisTyrrLeu 260
Db 944 CAGCTGCTGCTCTCCACCATCTGCTGATGAGAAACATCGAGAGAGAAACCATCTTCTTA 1003
Qy 261 AspLeuAlaGlyIleGluAsnTyrrThrSerGlnPheGlyProGlySerProSerValAla 280
Db 1004 GATCTCCCGGAGTAAAGAACTACACGCTCCCAATTTGGGCTGCTCCCTTCCTGGGCC 1063
Qy 281 GlnLysSerGluLeuProProArgThrSerAspProThrArgSerArgSerHisGluPro 300
Db 1064 CAGAAAGTCAGAAACGCCCCCGACCTCCATCCATCTCATCTCTCCATGAGACCG 1123
Qy 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db 1124 GAGACCATCCACATCCACACCGAAAGCCCAAGCGGTGACCGGCTCTCTTCACTTC 1183
Qy 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
Db 1184 CTTGACACCCCAATCGCCAGAGTCTCAGAGCTCCAGCAAGGCTCCGGGACACCAAGAC 1243
Qy 341 GlySerLysHisPheValArgSerProLysValAlaGlnGlyLysSerValGlyValGlyHis 360
Db 1244 GGGAGCAGGCACTTTGAGAGGTCTCCCAAGGCCCAAGGCAAGATGTGGGTGGGCGAC 1303
Qy 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1304 GTGGCCAGAGGGGCAAGAAACAGCCCTTGGGACCCCGCATCTCTGCGGTCTCCCC 1363
Qy 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1364 TCCGCCCACTGGCTGCAGCGCGGCTCTCTCCCTCCATGAGCCCGCTGGGACACAG 1423
Qy 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1424 AAGACACAGGCAACGAGCCAGAGAGCCAGCAGGCTGCGGGGCTGCAGGCACTG 1483
Qy 421 AlaSerGlyGlyProValLeuGlyArgGluHisIleLeuArgGlyLysProAlaLeuValVal 440
Db 1484 GCTTCAAGTGGCTCTGTCTGGGCGGAGCACTTGGGAGAGCTGCGGCTTGTGTGTG 1543
Qy 441 TyrgluSerGlnAlaGlyGlnProValGlnArgHisGluHisIleHisIleGluHis 460

Db 1544 TATGAGAGCAGGCGGGGACCGGCTCCAGAGACATGAGACCAACCACTGAACAT 1603
Qy 461 HisHisIleTyrrHisHisPheTyrglnThr 470
Db 1604 CACCACCATTAACACCACTTCTACAGACA 1633
RESULT 4
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902
Alignment Scores:
Pred. No.: 1,3e-185 Length: 2379
Score: 196.00 Matches: 196
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.70% Indels: 0
Gaps: 0
US-09-993-966-7 (1-470) x US-09-764-891-5902 (1-2379)
Qy 275 GlySerProSerValAlaGlnLysSerGluLeuProProArgThrSerAspProThrArg 294
Db 840 GGCCTCCCTCCGTTGGGCCAGAGTCAAGACTGCCCCCGCACCTCCATCCACTGCA 889
Qy 295 SerArgSerHisGluProGlnAlaIleHisIleProHisArgGlySPROGlnGlyValAsp 314
Db 900 TCTCGCTCCCATAGCGGAGGCGGACCATCCACATCCACAGAAAGCCCAAGCGTGGAC 959
Qy 315 ProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGlnArg 334
Db 960 CCGGCTCTCTCACTTCTCTTGACACCCCAATGCGCAAGTCTCAGAGCTCCAGAACGG 1019
Qy 335 LeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysValAlaGlnLys 354
Db 1020 CTCGGGGGACCCAGAGCGGAGCAAGCACTTGTGAGTCCCCCAAGGCCAGGCGAAG 1079
Qy 355 SerValGlyValGlnHisValAlaArgGlyValaArgAsnLysProProLeuGlyProAla 374
Db 1080 AGTGTGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAAGCCCTCTGGGACCGGCC 1139
Qy 375 IleProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeu 394
Db 1140 ATCCCTGCGGTCTCCCTCTCCGCCACTTGCTGCGAGCCCGGCTCTCTCCCTCTCTTA 1199
Qy 395 AlaProLeuGlyHisLysLysLysHisLysValAlaLysGluSerGlnGlnGlyCysArg 414
Db 1200 GCCCGCTCGGAGCAAGAGGACCAAGCAACGAGCCAGAGAGCCAGGAGCTGCGCG 1259
Qy 415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlnGlyArgGlnHisIleAspArg 434
Db 1260 GGGCTGAGGACACCACTGGCTCAGGTGGCCCTGTCTGGGGGGAGGACCTGCGGGAG 1319
Qy 435 LeuProAlaLeuValValTyrgluSerGlnAlaGlyGlnProValGlnArgHisGluHis 454
Db 1320 CTGGCCCGCTTGGTGTGTGTGAGAGCCAGGCGGCGGAGCCGCTCCAGACATGAGCAC 1379
Qy 455 HisHisIleHisGluHisIleHisIleTyrrHisIlePheTyrglnThr 470

Db 1380 CACCACCATGAAATCATCACACCATTACCAACCACTTCTACAGACA 1427

RESULT 5

US-09-764-891-5904

Sequence 5904, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION data removed - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5904

LENGTH: 2379

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-891-5904

Alignment Scores:

Pred. No.:	1,3e-185	Length:	2379
Score:	196.00	Matches:	196
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.70%	Indels:	0
	10	Gaps:	0

US-09-993-966-7 (1-470) x US-09-764-891-5904 (1-2379)

Qy 275 G1ySerProSerVal1aG1nLySerG1uLeuProProA1gThrSerAenProThArg 294

Db 840 GGCCTCCCTTCCTGCGCCAGAACTCAGAACTGCCCCCTCCATCCCACTCGA 899

Qy 295 SerArgSerHisG1uProG1uA1a1eHis1eProHisArgLySPProG1nLyVal1Asp 314

Db 900 TCTGCTCCATAGCGGAAGCATTCATCCACACCAAGCCCAAGGGGTGGAC 959

Qy 315 Pro1aSerPheHisPheLeuAspThrPro1eAlaLySValSerG1uLeuG1nArg 334

Db 960 CCGGCTCCTTCCACTTCTTGACACCCCAATGCCAAGCTCAGAGCTCCAGCAACGG 1019

Qy 335 LeuArgG1yThrG1nAspG1ySerLyHisPheVal1ArgSerProLyVal1aG1nLyLyS 354

Db 1020 CTCGGGGGCAACCGAGAGGAGCAAGCACTTGTGAGTCCCCCAAGGCCAGGGGCAAG 1079

Qy 355 SerValG1yValG1yHisVal1aArgG1yVal1aArgAsnLySPProLeuG1yProAla 374

Db 1080 AGTGTGGGTGGGCGACGTTGGCCAGAGGGGCAAGAAACAAGCCCCCTTGGAGCCGCC 1139

Qy 375 11ePro1aValSerProSerAlaHis1eVal1a1aSerProAlaLeuProSerLeu 394

Db 1140 ATCCCTGGGTGTCCTCCCTCGGCCCACTGCTGCCAGCCGGCTCTCTCCCTCTCTA 1199

Qy 395 AlaProLeuG1yHis1yLyHis1yLyHis1eArg1aLySValSerG1nG1yCysArg 414

Db 1200 GCCCCCTCGGGGCAAGAGCAACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1259

Qy 415 G1yLeuG1n1a1aProLeu1aSerG1yG1yProVal1eug1yArgG1yHis1eAlaArgG1u 434

Db 1260 GGCTGTAGGACCACTGGCTCAGTGGCTCTCTCTGGGGGGGAGCACTGGGGAG 1319

Qy 435 LeuPro1aLeuVal1yArg1ySerG1n1aG1yInProVal1G1nArgHis1eG1uHis 454

Db 1320 CTGGCGGCTTGGGTGATGAGCAAGCCAGCCGGGCAAGCCAGCCAGCCAGCCAGCCAG 1379

Qy 455 HisHisHis1eG1uHis1eHis1eHis1eHis1eHis1eHis1eHis1eHis1eHis1e 470

Db 1380 CACCACCATGAAATCATCACACCATTACCAACCACTTCTACAGACA 1427

RESULT 6

US-10-087-192-178

Sequence 178, Application US/10087192

Publication No. US20020182586A1

GENERAL INFORMATION:

APPLICANT: Morris, David W.

TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR

FILE REFERENCE: 52945200122

CURRENT APPLICATION NUMBER: US/10/087,192

PRIOR FILING DATE: 2002-03-01

PRIOR APPLICATION NUMBER: US 09/747,377

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: US 09/798,586

PRIOR FILING DATE: 2001-03-02

NUMBER OF SEQ ID NOS: 2059

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 178

LENGTH: 106315

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(106315)

OTHER INFORMATION: n = A,T,C or G

US-10-087-192-178

Alignment Scores:

Pred. No.:	3,29e-184	Length:	106315
Score:	196.00	Matches:	196
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.70%	Indels:	0
	13	Gaps:	0

US-09-993-966-7 (1-470) x US-10-087-192-178 (1-106315)

Qy 275 G1ySerProSerVal1aG1nLySerG1uLeuProProA1gThrSerAenProThArg 294

Db 95728 GGCCTCCCTTCCTGCGCCAGAACTCAGAACTGCCCCCTCCATCCCACTCGA 95787

Qy 295 SerArgSerHisG1uProG1uA1a1eHis1eProHisArgLySPProG1nLyVal1Asp 314

Db 95788 TCTGCTCCATAGCGGAAGCATTCATCCACACCAAGCCCAAGGGGTGGAC 95847

Qy 315 Pro1aSerPheHisPheLeuAspThrPro1eAlaLySValSerG1uLeuG1nArg 334

Db 95848 CCGGCTCCTTCCACTTCTTGACACCCCAATGCCAAGCTCAGAGCTCCAGCAACGG 95907

Qy 335 LeuArgG1yThrG1nAspG1ySerLyHisPheVal1ArgSerProLyVal1aG1nLyLyS 354

Db 95908 CTCGGGGGCAACCGAGAGGAGCAAGCACTTGTGAGTCCCCCAAGGCCAGGGGCAAG 95967

Qy 355 SerValG1yValG1yHisVal1aArgG1yVal1aArgAsnLySPProLeuG1yProAla 374

Db 95968 AGTGTGGGTGGGCGACGTTGGCCAGAGGGGCAAGAAACAAGCCCCCTTGGAGCCGCC 96027

Qy 375 11ePro1aValSerProSerAlaHis1eVal1a1aSerProAlaLeuProSerLeu 394

Db 96028 ATCCCTGGGTGTCCTCCCTCGGCCCACTGCTGCCAGCCGGCTCTCTCCCTCTCTA 96087

Qy 395 AlaProLeuG1yHis1yLyHis1yLyHis1eArg1aLySValSerG1nG1yCysArg 414

Db 96088 GCCCCCTCGGGGCAAGAGCAACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 96147

Qy 415 G1yLeuG1n1a1aProLeu1aSerG1yG1yProVal1eug1yArgG1yHis1eAlaArgG1u 434

Db 96148 GGCTGTAGGACCACTGGCTCAGTGGCTCTCTCTGGGGGGGAGCACTGGGGAG 96207

Qy 435 LeuPro1aLeuVal1yArg1ySerG1n1aG1yInProVal1G1nArgHis1eG1uHis 454

Db 96208 CTGGCGGCTTGGGTGATGAGCAAGCCAGCCGGGCAAGCCAGCCAGCCAGCCAGCCAG 96267

Qy 455 HisHisHis1eG1uHis1eHis1eHis1eHis1eHis1eHis1eHis1eHis1eHis1e 470

Db 96268 CACCACCATGAAATCAACACCATTCACACCATCTTACAGACACA 96315
RESULT 7
US-09-993-966-21
; Sequence 21, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHMAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/028073
; CURRENT APPLICATION NUMBER: US/09/993,966
; PRIOR FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,864
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-21
Alignment Scores:
Pred. No.: 3,97e-185 Length: 590
Score: 195.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.49% Indels: 0
DB: 10 Gaps: 0
US-09-993-966-7 (1-470) x US-09-993-966-21 (1-590)
Qy 276 SerProSerValAlaGlnLysSerGluLeuProProAlaGlnThrSerAnProThrArgSer 295
Db 3 TCCCTTCCTGGGCGGAGAGATCAAGACTGCCCGGACCTCCATCCCATCTCGAATCT 62
Qy 296 ArgSerHisGluProGluAlaIleHisIleProHisArgLysProGlnGlyValAspPro 315
Db 63 CGCTCCCATGAGCGGAGAGCATCATCCCAACGGAAGGCCCAAGGGTGGACCG 122
Qy 316 AlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGlnArgLeu 335
Db 123 GCCCTCTTCCACTTCTTGAACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGCTC 182
Qy 336 ArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysAlaGlnGlyLysSer 355
Db 183 CGGGGACCCGAGCGGAGAGCAAGCACTTGTGTGAGTCCCAAGGCCCAAGGCAAGGT 242
Qy 356 ValGlyValGlnHisValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIle 375
Db 243 GTGGGTGTGGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGCCGGCATC 302
Qy 376 ProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAla 395
Db 303 CCTCCGGGTGTCCCTCCCGCCCACTGCTGCCAGCCCGGCTCTCCCTCCCTCGCC 362
Qy 396 ProLeuGlyHisLysLysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGly 415
Db 363 CCCCCTGGGCAACAAGAGCAACAGCAACGAGCAAGAGCCAGAGGCTGCGGGAGC 422
Qy 416 LeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlyLysLysLeuArgLys 435
Db 423 CTGAGAGCAACCACTGGCTCAGTGGCTCTCTCTGGGGCGGAGCACTCGGGAGGTG 482
Qy 436 ProAlaLeuValValIlyrGluSerGlnAlaGlyLysProValGlnArgHisGlnHis 455
Db 483 CCGGCTTTGTGTGTGTGTGAGAGCCAGGCCGGGAGCGGCTCCAGAGCATGAGCACAC 542

Qy 456 HisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 543 CACCACCATGAAATCAACACCATTCACACCATCTTACAGACACA 587
RESULT 8
US-09-918-995-8686
; Sequence 8686, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8686
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(427)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8686
Alignment Scores:
Pred. No.: 2,92e-128 Length: 427
Score: 138.00 Matches: 138
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.36% Indels: 0
DB: 10 Gaps: 0
US-09-993-966-7 (1-470) x US-09-918-995-8686 (1-427)
Qy 285 LeuProProAlaGlnThrSerAnProThrArgSerHisGluProGluAlaIleHis 304
Db 4 CTCGCCCCCGGACCTCCATCCCATCTCGCTCCCATAGCGGAGCCGAAAGCCATTCAC 63
Qy 305 IleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeuAspThrPro 324
Db 64 ATCCCAACCGAAGAGCCCAAGGCGTGGAGCCGGCTCTTCCATCTCTTGACACCCCA 123
Qy 325 IleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAspGlySerLysHis 344
Db 124 ATGGCCAAAGTTCAGAGCTCCAGCAACGGCTCCGGGGCACCCAGGACGGAGCAAGCAC 183
Qy 345 PheValArgSerProLysAlaGlnGlyLysSerValGlyValGlnHisValAlaArgGly 364
Db 184 TTTGTAGGTCTCCCAAGGCGGAGAGAGTGTGGTGTGGGCCACGTGGCCAGAGGG 243
Qy 365 AlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSerAlaHisLeu 384
Db 244 GCAGAAACAAGCCCTCTGGAGCCGCCATCTCTCGGTGTCTCCCTCGGCCACCTG 303
Qy 385 AlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysLysHis 404
Db 304 GTGGGAGCCCGGCGCTCTCCCTCTCTGAGCCCTCTGGGAGCAAGAGCAAGAGCAC 363
Qy 405 ArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAlaSer 422
Db 364 CGAGCCAAAGAGGCAACAAGGGCTGCGGGGCGCTGCAACACACATCGGCTCA 417
RESULT 9
US-09-993-966-18
; Sequence 18, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:


```

?      APPLICANT: ROHMAN, MICHAEL
?      TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
?      TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
?      FILE REFERENCE: 014024/0280733
?      CURRENT APPLICATION NUMBER: US/09/993,966
?      CURRENT FILING DATE: 2001-11-27
?      PRIOR APPLICATION NUMBER: 60/252,884
?      PRIOR FILING DATE: 2000-11-27
?      PRIOR APPLICATION NUMBER: 60/291,109
?      PRIOR FILING DATE: 2001-05-16
?      PRIOR APPLICATION NUMBER: 60/325,571
?      PRIOR FILING DATE: 2001-10-01
?      NUMBER OF SEQ ID NOS: 26
?      SOFTWARE: PatentIn Ver. 2.1
?      SEQ ID NO: 18
?      LENGTH: 148
?      TYPE: DNA
?      ORGANISM: Homo sapiens
?      US-09-993-966-18

```

Pred. No.:	1.13e-39	Length:	148
Score:	49.00	Matches:	49
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	10.43%	Indels:	0
DB:	10	Gaps:	0
US-09-993-966-7 (1-470)	x	US-09-993-966-18 (1-148)	

[illegible]

```

RESULT 10
US-09-730-989-1
: Sequence 1, Application US/09730989
: Patent No. US20020061552A1
: GENERAL INFORMATION:
: APPLICANT: Yan, Dong
: APPLICANT: Williams, Lewis T.
: FILE OF INVENTION: MAMMALIAN DISCOVERED-ASSOCIATED PROTEINS
: FILE REFERENCE: PP-01657.002 / 200130.518
: CURRENT APPLICATION NUMBER: US/09/730,989
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 1401
: TYPE: DNA
: ORGANISM: Mus musculus
US-09-730-989-1

```

Alignment Scores:	
Pred. No.:	7,62e-39
Score:	49.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	10.43%
DB:	9
	Gaps: 0
	Length: 1401
	Matches: 49
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0

US-09-993-966-7 (1-470) x US-09-730-989-1 (1-1401)

131 Glu13AspSerArgGlnIuTrpThrPheThrLeuTyraaPheAspAsnGlyLys 150

Db	391	GAGGAGGACAGCGGGCAAGATGACATTTCACCTCATATGACTTCGACAAACAATGGCAAA	450
QY	151	ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValValAlaPseSer	170
Db	451	GTGACCCGCGAGACATTCACACGCTTCGCGCATACATCATGAAATGGTGACTCCTCT	510
QY	171	ValAlaMetIleSerProThrSerSerLys	179
Db	511	GTGAACCATTCCTCCCACTCAACGACAG	537

```

? RESULT 11
? US-09-993-966-6
? Sequence 6, Application US/09993966
? Publication No. US20030186232A1
? GENERAL INFORMATION:
? APPLICANT: ROHAN, MICHAEL
? TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NMD PROTEIN
? TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
? FILE REFERENCE: 014024/028073
? CURRENT APPLICATION NUMBER: US/09/993,966
? CURRENT FILING DATE: 2001-11-27
? PRIOR APPLICATION NUMBER: 60/252,884
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/291,109
? PRIOR FILING DATE: 2001-05-16
? PRIOR APPLICATION NUMBER: 60/335,571
? PRIOR FILING DATE: 2001-10-01
? NUMBER OF SEQ ID NOS: 26
? SOFTWARE: PatentIn Ver. 2.1

```

```

; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-6

Alignment Scores:
Pred. No.:      7,62e-39
Score:          49.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:    10.43%
DB:             10

Length: 1401
Matches: 49
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```

US-09-993-966-7 (1-470) x US-09-993-966-6 (1-1401)

QY	131	GIUGIUAAPSErArGIGlInUTPThrPherTrLeUryAspPhehApSAnEnGlyLyS	150
Db	391	GAGGAGACAGCGGCAGAGAGTGACTTCACTATATGACTTCACAACATGGCANA	450
QY	151	VAlTrArAgGlUAAPLleThrSerLeuLeuHsrTrLleTyGLuValValApsSerSer	170
Db	451	GTGACCCTGTGAGCATTAACGACTGTGTCATTACATCTAAGAAGTGTTGACTCTCT	510
QY	171	ValAenHisserProThrSerSerLyS	179
Db	511	GTGAACCATTCCTCCACATCAAGCAG	537

RESULT 12
US-09-993-966-2
; Sequence 2, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:

?
? APPLICANT: ROHANI, MICHAEL
? TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN
? TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
? FILE REFERENCE: 014024/0280733
? CURRENT APPLICATION NUMBER: US/09/993, 966
? CURRENT FILING DATE: 2001-11-27
? PRIOR APPLICATION NUMBER: 60/252,884
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/291,109
? PRIOR FILING DATE: 2001-05-16
? PRIOR APPLICATION NUMBER: 60/325,571
?

```
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-2

Alignment Scores:
Pred. No.: 7.69e-39      Length: 1416
Score: 49.00           Matches: 49
Percent Similarity: 100.00%  Conservat: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 10.43%      Indels: 0
DB: 10                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-993-966-2 (1-1416)

QY 131 GluGluAspSerArgGlnGluTrpThrPheThrLeuTyrAspPheAspAsnGlyLys 150
Db 391 GAGGAGGACGACCGCGAAGAGTGACTTCACTTATATGACTTCGACCAATGGCAAA 450
QY 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValValAspSerSer 170
Db 451 GTGACCGGTGAGGACATTCACGCTTGTCATCCATCTATGAAAGTGTTGACTCTCT 510
QY 171 ValAsnHisSerProThrSerSerLys 179
Db 511 GTGAACCATTCCTCCACATCAAGCAAG 537

RESULT 13
US-10-087-192-176
; Sequence 176, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176
; LENGTH: 2807
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-176

Alignment Scores:
Pred. No.: 1.38e-38      Length: 2807
Score: 49.00           Matches: 49
Percent Similarity: 100.00%  Conservat: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 10.43%      Indels: 0
DB: 13                  Gaps: 0

US-09-993-966-7 (1-470) x US-10-087-192-176 (1-2807)

QY 131 GluGluAspSerArgGlnGluTrpThrPheThrLeuTyrAspPheAspAsnGlyLys 150
Db 333 GAGGAGGACGACCGCGAAGAGTGACTTCACTTATATGACTTCGACCAATGGCAAA 392
QY 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValValAspSerSer 170
Db 393 GTGACCGGTGAGGACATTCACGCTTGTCATCCATCTATGAAAGTGTTGACTCTCT 452
```

```
QY 171 ValAsnHisSerProThrSerSerLys 179
Db 453 GTGAACCATTCCTCCACATCAAGCAAG 479

RESULT 14
US-10-016-157A-126
; Sequence 126, Application US/10016157A
; Publication No. US20020192220A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Ghosh, Malavika
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Prot
; FILE REFERENCE: DEX-0253
; CURRENT APPLICATION NUMBER: US/10/016,157A
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,717
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 250
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 126
; LENGTH: 1950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-157A-126

Alignment Scores:
Pred. No.: 1.01e-35      Length: 1950
Score: 46.00           Matches: 46
Percent Similarity: 100.00%  Conservat: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 9.79%      Indels: 0
DB: 13                  Gaps: 0

US-09-993-966-7 (1-470) x US-10-016-157A-126 (1-1950)

QY 20 GlYAspSerPheAlaValSerAlaIleTrrPAlaArgLysGlyTlleglGluTrpIlegly 39
Db 838 GGTGACGCTTCCTCCGTTGACGCTGCTGGCTCGAAGGCAATCGAGAGTGATCGG 897
QY 40 ArgGlnArgCyseProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
Db 898 AGACAGGCGTGGCCGGGGCGGTGTCTCGGAGCCCGACAGCTGCGGTGGCGGACCATTA 957
QY 60 GlYArgSerThrArgGlu 65
Db 958 GGCCGAAGCACCCGGGAA 975

RESULT 15
US-10-029-386-7990
; Sequence 7990, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7990
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
```

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
 ; OTHER INFORMATION: SWISSPROT HIT: Q23977, EVALUE 7.70e-02
 ; OTHER INFORMATION: NT HIT: g14916432, EVALUE 9.00e-71
 ; OTHER INFORMATION: EST_HUMAN HIT: BF920913.1, EVALUE 3.00e-68
 US-10-029-386-7990

Alignment Scores:
 Pred. No.: 3.68e-35 Length: 598
 Score: 45.00 Matches: 45
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.57% Indels: 0
 DB: 15 Gaps: 0

US-09-993-966-7 (1-470) x US-10-029-386-7990 (1-598)

QY	20	GIYASPSERPhEALaValSerAlaAlaTTPAlaArgLySGIYIleGIuGIuTPiIleGIY	39
Db	229	GGTGACAGCTTCGCCGTGAGCGCTGCGGCTCGGAAGGCATCGAGAGTGATCGGG	288
QY	40	ArgGlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle	59
Db	289	AGACAGCGCTGCCCGGCGGTGTCTCGAGACCCCGACAGCTGCGGTGCGGSCACCATTA	348
QY	60	GIYArgSerThrArg	64
Db	349	GGCCGAGCACCCCGG	363

Search completed: December 30, 2004, 14:01:27
 Job time: 995 secs

THIS PAGE IS BLANK

Result No.	Score	Query Match	Length	DB	ID	Description
1	470	100.0	1859	10	AA151511	AA151511 Human Ntk
2	383	81.5	1438	3	AA663895	AA663895 DNA encod
3	196	41.7	2379	4	AA103214	AA103214 Human rep
4	196	41.7	2379	4	AA103216	AA103216 Human rep
5	195	41.5	550	10	AA151525	AA151525 Human Ntk
6	138	29.4	427	9	ACH21474	ACH21474 Human adu

7	107	22.8	1561	5	AA6S2082	DNA	encod
8	72	15.3	1743	5	AA78752	DNA	encod
9	72	15.3	1743	5	AA692081	DNA	encod
10	72	15.3	1743	10	ADG31866	Human	nov
11	49	10.4	148	10	AAJ51522	Human	NK6
12	49	10.4	1401	10	AAH51512	Mouse	NK6
13	49	10.4	1416	4	AAH28343	Nucleotid	
14	46	9.8	1950	6	AB073820	Human	col
15	45	9.6	588	12	ACH74795	Human	ger
16	45	9.6	1731	3	AAAC3924	DNA	encod
17	44	9.4	133	12	ACH88495	Human	ger
18	44	9.4	134	10	AAJ51518	Human	NK6
19	42	8.9	128	10	AAJ51524	Human	NK6
20	36	7.7	639	8	ACD05637	CDNA	enco
21	22	7.4	107	10	AAJ51520	Human	NK6
22	22	6.8	96	10	AAJ51521	Human	NK6
23	29	6.2	554	12	ACH74420	Human	ger
24	28	6.0	179	12	ACH88120	Human	ger
25	27	5.7	85	10	AAJ51523	Human	NK6
26	22	4.7	67	10	AAJ51519	Human	NK6
27	17	3.6	1285	3	AAAC3926	DNA	encod
28	17	3.6	1307	3	AAAC3927	DNA	encod
29	17	3.6	1817	6	ABK49445	Human	CDN
30	29	3.6	1885	10	ADP82455	Leukemia	
31	17	3.6	1959	12	AD224400	Human	scot
32	17	3.6	1966	4	AA677117	CDNA	enco
33	17	3.6	1966	10	ADBP93295	Human	CDN
34	13	3.0	551	10	ADP81081	Leukemia	
35	14	2.8	2723	4	AAK75944	Human	imm
36	10	2.3	34	10	AAJ51517	Human	NK6
37	10	2.1	574	4	AAI10301	Human	NK6
38	10	2.1	574	4	ABAA5194	Human	foe
39	10	2.1	574	4	AAJ15552	Human	foe
40	10	2.1	574	4	ABAA1762	Probe	#23
41	10	2.1	574	4	AAK25674	Human	bon
42	10	2.1	574	4	AAK00231	Human	bra
43	10	2.1	574	4	AB825257	Human	liv
44	10	2.1	574	5	AAI00237	Probe	#22
45	10	2.1	574	6	AB800249	Human	gen

ALIGNMENTS

```

RESULT 1
AAL51511
ID AAL51511 standard; DNA; 1859 BP

```

Human Nkd (hNkd) protein coding sequence.

KW Human; gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;
 KW colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.

OS Homo sapiens

FH	Key	Location/Qualifiers
20	100	2010

FT	CDS	224.	.1636
----	-----	------	-------

```

FT      /product= "Human Nkd protein"

```

PN WO200292832-A2

PD 21-NOV-2002

PF 27-NOV-2001; 2001WO-US044092
 vv

PR 27-NOV-2000; 2000US-0252884P.
PR 15-MAY-2001; 2001US-0281109P.
PR 15-MAY-2001; 2001US-0281109P.

PR 01-OCT-2001; 2001US-0325571P.
XX XX

PA (CHIR) CHIRON CORP.
 XX
 PI Rohan M, Chan V, Yan D;
 XX
 DR WPI, 2003-129303/12.
 DR P-PSDB; AAO16389.
 XX
 XX New human and non-human primate homologues of Nkd protein, and Nkd genes,
 PT useful for treating cancer involving aberrant Wnt signaling, e.g. colon
 PT cancer, head and neck cancer, ovarian cancer, or breast cancer.
 XX
 XX Claim 4; Fig 3; 99pp; English.
 XX
 CC The invention comprises the amino acid and coding sequence of the human
 CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
 CC and protein sequences are useful for the treatment of cancer involving
 CC aberrant Wnt signaling (e.g. colon cancer, head and neck cancer, ovarian
 CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
 CC protein of the invention
 XX
 SQ Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 1859
 Score: 470.00 Matches: 470
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-993-966-7 (1-470) x AML5151 (1-1859)
 QY 1 MetGlyIysLeuHisSerIyProAlaAlaValCylsYsarGdluSerProGluGly 20
 Db 224 ATGGGGAAACTTCACTCCAAAGCCGCGCCGTGTGCAAGCCAGAGAGAGCCGAAAGGT 283
 QY 21 AspSerPheAlaValSerAlaAlaTrrAlaArgIySglYlleglUGlUtrPllleGlyArg 40
 Db 284 GACAGCTTCGCCGCTGAGCGCTGCTGGAGCTCGAAGGGCATCGAGAGTGGATCGGAGGA 343
 QY 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
 Db 344 CAGGCGTCCCGGCGGCTGTGGAGCCCGGACAGCTGCGGTTGGCGGCGGACCATGGGC 403
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
 Db 404 CGAAGCACCCGGGAGCTGTGGGCGAGCTGTGAGACACGCTCAGCGAGAGAGAGAG 463
 QY 81 AspAspPheArgLeuGluValAlaAlaLeuProProGluThrArgPglYleuGlySerGly 100
 Db 464 GAGGACTTTGGGCTGGAGTGGCTGCTCTCTTGAAGACTGACGGGCTGGGCGAGGGA 523
 QY 101 AspGluIySlyMetGluArgValSerGluProCysProGlySerIySlySglNleuLys 120
 Db 524 GATGAGAAAGAGATGAGAGAGTGGAGGACCTCGCCAGGCTCCAGAAAGCACTGAG 583
 QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
 Db 584 TTTGAAAGGCTCCAGTGGAGCTGTCCATGAGAGAGACAGCCGCGAGAGTGGACCTTC 643
 QY 141 ThrLeuTrpAspPheAspAsnAngIyValAlaThrArgGluAspPllleThSerLeuLeu 160
 Db 644 ACCCTGATGACTTTCACACACGAGGAGGTCAACCGAGAGGACATCACAGCTTGGTG 703
 QY 161 HisThrIleTrpGluValAlaAspSerSerValAsnHisSerProThrSerSerIyMet 180
 Db 704 CACACCATCATATGAGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATG 763
 QY 181 LeuArgValIySleuThrValAlaProAspGlySerGlnSerIySarGserValleuVal 200
 Db 764 CTGGGGGTAAAGCTCACCGTGGCCCGATGGACGACAGACAGAGAGGCTGCTTGTCTC 823
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIySProThrGluAspLeu 220

Db 824 AATCAGGCTGACCTGCAGAGCGAGGCGCCCGAGCAGAGACCAAGCCCACTGAGGACTG 883
 QY 221 ArgSerTrpGluIySlySglNleuArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 Db 884 CGGAGCTGGGAGAAAGAGCAGCGAGCGCCCGCTCAGGTTCACGGGTGACACCGGCTGGAG 943
 QY 241 GlnSerGlyCysTrpHisSlyCysValAspGluAsnIleGluArgArgAsnHisTrpLeu 260
 Db 944 CAGTCTGGCTGTACCACTTGGCTGTGATGAGAACTTCAGAGAGAAACCACTACTTGA 1003
 QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
 Db 1004 GATCTGCCGGGATAGAAACTACAGCTCCCAATTTGGGCTGCTCCCTTCGTTGAGCC 1063
 QY 281 GlnIySserGluLeuProProArgThrSerAsnProThrArgSerArgIySglPro 300
 Db 1064 CAGAACTCAGAACTGCCCCCGACCTCCAAATCCCATGATCTCCCTCCCAAGACCCG 1123
 QY 301 GluAlaIleHisIleProHisArgIySProGlnGlyValAspProAlaSerPheHisPhe 320
 Db 1124 GAAGCATCCACATCCACACCGAAAGCCCAAGGCGTGGACCGGCTCTTCCACTTC 1183
 QY 321 LeuAspThrProIleAlaIySValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
 Db 1184 CTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAGAGGCTCGGGGACCCAGCAGAC 1243
 QY 341 GlySerIySlySlyPheValArgSerProIyAlaGlnGlyIySserValGlyValIyHis 360
 Db 1244 GGGAGCAAGCACTTGTGAGGTCCCGCAAGGCCAGGCGCAAGTGTGGGTGTGGGCAC 1303
 QY 361 ValAlaArgGlyAlaArgAsnIyProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 1304 GTGGCCAGAGGGGCAAAACAAAGCCCTCTTGGAACCCCATCTCGGCTGTCCCCC 1363
 QY 381 SerAlaHisLeuAlaIleSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 Db 1364 TCCGCCCACTGGGTGCACCGCGGCTCTCCCTCCCTGATGCCCTCCGCGCACAG 1423
 QY 401 LysHisLysHisArgAlaIySglNleuArgGlnGlnGlyCysArgIyLeuGlnAlaProLeu 420
 Db 1424 AAGCACAAGCAGCGAGCCAGAGAGCCAGAGGCTGCCGGGCTCGCAGGCGCACACTG 1483
 QY 421 AlaSerGlyIyProValIleuGlyArgGluHisIleuArgGluLeuProAlaLeuValVal 440
 Db 1484 GCTTCAGGTGCTCTGTGTGGGCGGAGACCTGCGGAGCTGCCGCTTGTGTGTG 1543
 QY 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisIleGluHisIleHisIleGluHis 460
 Db 1544 TATGAGAGCAGGCGCGGAGCGGTCAGAGACATGAGACACCAACCATGAACAT 1603
 QY 461 HisHisIleTrpHisIlePheTrpGlnThr 470
 Db 1604 CACCACCATTAACACCACTTCTACAGACA 1633
 RESULT 2
 AAA63925
 ID AAA63925 standard; DNA; 1438 BP.
 AC AAA63925;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE DNA encoding protein related to Drosophila naked cuticle polypeptide.
 XX Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signaling;
 KW gene therapy; Nkd defect; cancer; ss.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 8..1420
 FT CDS
 FT /*tag= a

FT /product= "naked cuticle polypeptide"

XX MO200049034-A1.

XX PD 24-AUG-2000 a date

XX PF 17-FEB-2000; 2000MO-US0004188.

XX 17-FEB-1999; 99US-0120646P

XX (STRD) UNIV LELAND STANFORD JUNIOR.

XX Scott M, Zeng W, Wharton K;

XX WPI; 2000-571967/53.

XX P-PSDB; AAB08216.

XX An isolated nucleic acid molecule useful for analyzing (genetic

PT predileposition to) a disease state and for therapeutic purposes e.g.

PT treatment of cancer comprises a sequence encoding a naked cuticle

PT protein.

XX Claim 3; Page 46-48; 58pp; English.

XX The present sequence encodes a protein related to the Drosophila Nkd

CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity

CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide

CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the

CC regulation of Wnt signal potency, duration or distribution. The Nkd

CC polynucleotides can be used for identifying homologous or related

CC proteins, to modulate the expression or function of Nkd polypeptides, and

CC in studying associated physiological pathways. Nkd polynucleotides can

CC also be used in gene therapy to treat disorders associated with Nkd

CC defects. They may also be used for therapeutic purposes e.g. treatment of

CC cancer

XX SQ Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	1438
Score:	383.00	Matches:	469
Percent Similarity:	99.58%	Conservative:	0
Best Local Similarity:	99.58%	Mismatches:	1
Query Match:	81.49%	Indels:	2
		Gaps:	0

US-09-993-966-7 (1-470) x AAA63925 (1-1438)

QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLysArgArgLysSerProGlyGly 20

DB 8 ATGGGGAACTTCCTCAAGCCCGCGCGTGTGCAGAGCCGAGAGAGCCCGAAGCT 67

QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyIleGlyGlyIleGlyArg 40

DB 68 GACAGCTTCCCGTGAAGCGCTGCTGGCTCGAAGGAGCATCGAGAGTGGATCGGAGA 127

QY 41 GlnArgCysArgProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60

DB 128 CACGCGCTCCCGGGCGGTGTCTCGGAGACCCGACAGCTGGCGTTGGGGGACCATGGC 187

QY 61 ArgSerThrArgGlnLeuValGlyAspValLeuArgAspThrLeuSerGlnGlnGlu 80

DB 188 CGAAGCACCAGGAGCTCGTGGGAGAGTGTGAGACACGCTCACAGAGAGAGAG 247

QY 81 AspAspPheArgLeuGlnValAlaLeuProProGlyLysThrAspGlyLeuGlySerGly 100

DB 248 GAGGACTTTGGCTGGAGAGTGGCTTCTCTCTAGAGAGACTGACGGGCTGGAGAGG 307

QY 101 AspGlnLysLysMetGlnArgValSerGlyProCysArgProGlySerLysGlnLeuLys 120

DB 308 GATGAGAAAGAGAGAGAGAGTGAAGAGAGCCCTGCCAGGCTCCAGAGAGAGAGAG 367

QY 121 PheGlnGlnLeuGlnCysAspValSerMetGlnGlnLysAspSerArgGlnGlnThrPhe 140

DB 368 TTGAGAGCTCCAGTCCAGTGTCCATGAGAGAGAGCCGCGAGAGTGAACCTTC 427

QY 141 ThrLeuTyraAspPheAspAsnGlyLysValThrArgLysAspIleThrSerLeuLeu 160

DB 428 ACCCTGTATGACTTTGACCAACAGGAGAGTACCCGAGAGAGATCACAGCTTGTGTG 487

QY 161 HisThrIleTyrglnValValAspSerSerValAsnHisSerProThrSerSerLysMet 180

DB 488 CACACCATCTATGAGTGTGTGACTCTCTGTCAACCATCCCAATCATCAGCAAGATG 547

QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200

DB 548 CTGGGGTAAAGCTCACCGTGGCCCGATGGACGACCAAGAGAGAGAGAGGCTCTGTTC 607

QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnThrLysProThrGlnAspLeu 220

DB 608 AATCAGGCTGACCTGCAGAGCGGAGGCCCGGAGCAGACAGACCAAGCCCATGAGAGACTG 667

QY 221 ArgSerTrpGlnLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240

DB 668 CGGAGCTGGAGAGAGAGAGAGAGCGGCCCTCAGGTTCCAGGAGTGAAGAGCCGCTGAG 727

QY 241 GlnSerGlyCysTyraHisLysCysValAspGlnAsnIleGlnLysArgAsnHisTyrlau 260

DB 728 CAGCTGGCTGTACCACTTCGTTGATGATGAGAACATCGAGAGAGAGAAACCTACTTGA 787

QY 261 AspLeuAlaGlyIleGlnAsnTyraSerGlnPheGlyProGlySerProSerValAla 280

DB 788 GATTCGCCCGGATGAAATCTACACTCCCATTTGGGCTGTGCTCCCTTCCGTGGCC 847

QY 281 GlnLysSerGlnLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPr 300

DB 848 CAGAACTCAAGACTGCCCCCCCCCAGACCTTCAATCCCACTGATCTGCTCCCATGAGCC 906

QY 300 OGAlaAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisph 320

DB 907 GGAAGCCATCCATCCACACCGAAGAGCCCAAGAGGAGTGAACCCGCTCTTCCACTT 966

QY 320 eLeuAspThrProIleAlaLysValSerGlnLeuGlnArgLeuArgLysThrGlnAs 340

DB 967 CTTTGAACCCCAATGCCCAAGGCTCTAGAGCTTCAGCAAGCGCTCGGGGCACTCAAGA 1026

QY 340 PglySerLysHisIlePheValArgSerProLysValGlnGlyLysSerValGlyHis 360

DB 1027 CGGAGAGCAACACTTGTGTGAGTCCCCCAAGGCCAGGAGAGTGTGGGTGTGGGCCA 1086

QY 360 sValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPr 380

DB 1087 CGTGGCCAGAGGAGGCAAGAAACAGCCCTCTGGAGCCCGCATCTCTCGGTGTCCC 1146

QY 380 sSerAlaHisIleuAlaHisSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisly 400

DB 1147 CTCGCGCCACTGTGCTCCAGCCCGGCTCTCTCCCTCCCTAGCCCCCTCGGGCAAA 1206

QY 400 sLysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420

DB 1207 GAAAGCAACAGCAGGAGCCAGAGAGAGCCAGAGGCTGCGGGGAGCTGAGAGGACCACT 1266

QY 420 uAlaSerGlyGlyProValLeuGlnArgGlnHisIleuArgGlnLeuProAlaLeuVala 440

DB 1267 GGGCTCAAGTGGCCCTGTCTCTGGAGGAGAGACCTGCGGAGAGTGGCCCTGTGGTGT 1326

QY 440 lTyrgLysSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisGlnHis 460

DB 1327 GTATGAGAGCAAGCCGCGGAGCGCTCGGACAGACATGAGACACCAACCATGAAACA 1386

QY 460 HisHisHisHisTyraHisIlePheTyrglnThr 470

DB 1387 TCACCAACATTAACACCACTTCAACAGACA 1417

RESULT 3

AAI03214

ID AAL03214 standard; DNA: 2379 BP.
XX
AC AAL03214;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5902.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 08-NOV-2000; 2000US-0246652P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.

[illegible]

PR	08-SEP-2000	2000US-0231244F
PR	08-SEP-2000	2000US-0231143P
PR	08-SEP-2000	2000US-0231414P
PR	08-SEP-2000	2000US-0232080P
PR	08-SEP-2000	2000US-0232081P
PR	12-SEP-2000	2000US-0231968P
PR	14-SEP-2000	2000US-0232297P
PR	14-SEP-2000	2000US-0232398P
PR	14-SEP-2000	2000US-0232399P
PR	14-SEP-2000	2000US-0232400P
PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0233063P
PR	14-SEP-2000	2000US-0233064P
PR	14-SEP-2000	2000US-0233065P
PR	21-SEP-2000	2000US-0234423P
PR	21-SEP-2000	2000US-0234474P
PR	25-SEP-2000	2000US-0234977P
PR	25-SEP-2000	2000US-0234988P
PR	26-SEP-2000	2000US-0235484P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235835P
PR	29-SEP-2000	2000US-0236327P
PR	29-SEP-2000	2000US-0236367P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236369P
PR	29-SEP-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0236802P
PR	02-OCT-2000	2000US-0236823P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	02-OCT-2000	2000US-0237040P
PR	13-OCT-2000	2000US-0237935P
PR	13-OCT-2000	2000US-0239937P
PR	20-OCT-2000	2000US-0240960P
PR	20-OCT-2000	2000US-0241182P
PR	20-OCT-2000	2000US-0241785P
PR	20-OCT-2000	2000US-0241786P
PR	20-OCT-2000	2000US-0241787P
PR	20-OCT-2000	2000US-0241808P
PR	20-OCT-2000	2000US-0241809P
PR	01-NOV-2000	2000US-0244617P
PR	08-NOV-2000	2000US-0246474P
PR	08-NOV-2000	2000US-0246475P
PR	08-NOV-2000	2000US-0246476P
PR	08-NOV-2000	2000US-0246477P
PR	08-NOV-2000	2000US-0246478P
PR	08-NOV-2000	2000US-0246523P
PR	08-NOV-2000	2000US-0246524P
PR	08-NOV-2000	2000US-0246555P
PR	08-NOV-2000	2000US-0246556P
PR	08-NOV-2000	2000US-0246557P
PR	08-NOV-2000	2000US-0246558P
PR	08-NOV-2000	2000US-0246559P
PR	08-NOV-2000	2000US-0246609P
PR	08-NOV-2000	2000US-0246610P
PR	08-NOV-2000	2000US-0246611P
PR	08-NOV-2000	2000US-0246613P
PR	17-NOV-2000	2000US-0249207P
PR	17-NOV-2000	2000US-0249208P
PR	17-NOV-2000	2000US-0249209P
PR	17-NOV-2000	2000US-0249210P
PR	17-NOV-2000	2000US-0249211P
PR	17-NOV-2000	2000US-0249212P
PR	17-NOV-2000	2000US-0249213P
PR	17-NOV-2000	2000US-0249214P
PR	17-NOV-2000	2000US-0249215P
PR	17-NOV-2000	2000US-0249216P
PR	17-NOV-2000	2000US-0249217P
PR	17-NOV-2000	2000US-0249218P
PR	17-NOV-2000	2000US-0249244P
PR	17-NOV-2000	2000US-0249245P
PR	17-NOV-2000	2000US-0249246P

PR	17-NOV-2000;	2000US-0249255P.	
PR	17-NOV-2000;	2000US-0249297P.	
PR	17-NOV-2000;	2000US-0249299P.	
PR	17-NOV-2000;	2000US-0249300P.	
PR	01-DEC-2000;	2000US-0250160P.	
PR	01-DEC-2000;	2000US-0250391P.	
PR	05-DEC-2000;	2000US-0251030P.	
PR	05-DEC-2000;	2000US-0251988P.	
PR	05-DEC-2000;	2000US-0256719P.	
PR	06-DEC-2000;	2000US-0251479P.	
PR	08-DEC-2000;	2000US-0251868P.	
PR	08-DEC-2000;	2000US-0251869P.	
PR	08-DEC-2000;	2000US-0251989P.	
PR	08-DEC-2000;	2000US-0251990P.	
PR	11-DEC-2000;	2000US-0254097P.	
PR	05-JAN-2001;	2001US-0259678P.	
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI: 2001-465570/50.		
XX			
PT	Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.		
XX			
PS	Disclosure; SEQ ID NO 5904; 1297pp + Sequence Listing; English.		
XX			
CC	The present invention provides the protein and coding sequences of a		
CC	number of human reproductive system related antigens. These can be used		
CC	in the prevention and treatment of reproductive system disorders,		
CC	including cancer. The present sequence is a genomic sequence encoding a		
CC	protein of the invention		
XX			
SQ	Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	1.08e-179	Length:	2379
Score:	196.00	Matches:	196
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.70%	Indels:	0
DB:	4	Gaps:	0
US-09-993-966-7 (1-470) x AAL03216 (1-2379)			
QY	275	GLYSERPROSERVALAAGLNLYSERGIULEUPROPROARGTHRSEANPROTHARG	294
DB	840	GGCTCCCTTCGTTGCGCCACGAAGTCAGAACTGCCCCCGACCTCCATCCACTCGA	899
QY	295	SEATARGSERHISGLUPROGLUNLATLEHISILEPROHISARGLYSPROGLINGLYVALAEP	314
DB	900	TCTCCCTCCATGACCCGGAAGCCATCCATCCACACCCGAAAGCCCAAGCGCGAGC	959
QY	315	PROALASERPHENISAPHELEUAPRTHRPROILEALALYVALSERGIULEUNGINGNARG	334
DB	960	CCGGCTCCTTCCACTTCTCCTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGG	1011
QY	335	LEUARGLYTHRGLAASPGLYSERLYSHISPIEVALARGSERPROLYVALAGLNGLYLYS	354
DB	1020	CTCCCGGGACCCACGAGACGGGAGCACTTTGTGAGGTCCCCCAAGGCCACGAGGCAAG	107
QY	355	SERVALGIVALGIVHISVALAALAAARGGLVLAARGAENLVSPROBLEUGLYPROALA	374
DB	1080	AGTGTGGGTGGGGCAGCTGGCCAGAGGGGCAAGAAACMACCCCTCTGGGACCCGCC	113
QY	375	ILEPROALAVASERPROSERALAHISLEUALAALASERPROALALEUENUPROSERLEU	394
DB	1140	ATCCCTGGGGTCTCCCTCCCTCCGACCACTGGTGCACGGCGGCTCTCCCTCCCTCA	119
QY	395	ALAPROLEUGLYHISLYLYSHISLYSHIVARGALALYGLINSEGTINGLNGLYCYVARG	414

Dd	1200	GCCCCCCTCGGGGACCAAGAACACAGACCCGAGGCAGAGAGGCCACAGGCGTCCGGC	125
Oy	415	GLYLeuGIAlAPrOLEuLALSeRcLYGlYPVaLLeUGLYAArgLIuHISLeuArGLU	434
Dd	1260	GGCGTCGACGACCACTGGCTTCAGGTGGCCCTGCTCGGGGCGGAGACACTGGGGAG	1319
Oy	435	LeuPrOAlleuValTYrGLuSerGlnIaGLYgInPProVALGlnArGHISGLuHS	454
Dd	1320	CTGGCCCCGTTGGTGATAGAGCCAGGCGGGGAGCCGGTCCAGAGACATGAGCAC	1379
Oy	455	HISHisHSISGLuHISHSISHSISHTyrHISHSIPheTYrGLINthr	470
Dd	1380	CACCACCACTGAACATCACCATTAACAACACTTCTACAGACA	1427
RESULT 5			
ID	AAL51525	standard; DNA; 590 BP.	
XX	AAL51525;		
AC			
XX			
DT	24-APR-2003	(first entry)	
XX			
DE	Human Nkd (hNkd) gene exon 11 (coding region).		
XX			
KW	Human; de; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;		
KM	colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;		
OS	promoter.		
XX			
XX	Homo sapiens.		
PN	WO200292832-A2.		
XX			
PD	21-NOV-2002.		
XX			
PF	27-NOV-2001; 2001WO-US044092.		
PR	27-NOV-2000; 2000US-0252884P.		
PR	16-MAY-2001; 2001US-0291109P.		
PR	01-OCT-2001; 2001US-0325571P.		
XX			
PA	(CHIR) CHIRON CORP.		
PI	Rohan M, Chan V, Yan D;		
XX			
DR	WPI; 2003-129303/12.		
PT	New human and non-human primate homologues of Nkd protein, and Nkd genes,		
PT	useful for treating cancer involving aberrant Wnt signaling, e.g. colon		
FT	cancer, head and neck cancer, ovarian cancer, or breast cancer.		
XX			
PS	Example 3; Fig 7; 99p; English.		
XX			
CC	The invention comprises the amino acid and coding sequence of the human		
CC	Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA		
CC	and protein sequences are useful for the treatment of cancer involving		
CC	aberrant Wnt signaling (e.g. colon cancer, head and neck cancer, ovarian		
CC	cancer and breast cancer). The present DNA sequence represents exon 11		
CC	(coding region) from the human Nkd gene		
XX			
SQ	Sequence 590 BP; 119 A; 234 C; 159 G; 78 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	2.68e-179	Length:	590
Score:	195.00	Matches:	195
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.49%	Indels:	0
DB:	10	Gaps:	0
US-09-993-966-7 (1-470) x AAL51525 (1-590)			
Oy	276	SeRPoSRvAlAlAGlNySeRGluLeuPProFOrArghrSerAnProThArGer	295

Db		3 TCCCTTCCTCGGTGGCCCAAGATCAGAACTGCCCCCCCCGACCTCCATCCCATCCATCATCT	62
Qy		256 AAGSerHisAGLupProGluAlaIleHisIleProHisAATyLupProGlnGlyValAaspPro	315
Db		63 CGCTCCCATAGACCGGAAAGCCATCCACATCCACACCGAAAGCCCAAGGCGTGGACCCG	1222
Qy		316 AAsSerPheHisPheLeuAspThrProIleAlaValSerGluLeuGlnGlnGlyLeu	335
Db		123 GCTCTCTTCACCTCTCTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAAGGCTC	182
Qy		336 ATGGATYTHrGlnAAspGlySerTyHisHisPheValArgSerProValAaGlnGlyLysSer	355
Db		183 CGGGGACACCCAGAGACGGAGACAGCACTTTGTGAGTCCCCCAAGGCCCAAGGCAAGAGT	242
Qy		356 ValGlyValGlyHisValAlaAargGlyAlaArgAsnLysPheProLeuGlyProAlaIle	375
Db		243 GTGGGTGTGGGCGACGTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGGAGCCGCGCATC	302
Qy		376 ProAlaValSerProSerSerAlaHisLeuAlaAlaSerProAlaLeuPheSerLeuAla	395
Db		303 CCTGGGGGTGTCCCTCCCTCGGACCACTGGCTGGACGCCGAGCCCTCTCCCTCCAGCC	362
Qy		396 ProLeuGlnHisLeuLysLeuHisHisHisAATyAlaLysGluSerGlnGlnGlyCysArgGly	415
Db		363 CCCCTCGGGACAGAAAGACAAAGACCGAGCCCAAGAGACAGAGGAGCTCCGAGGC	422
Qy		416 LeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyAArgGluHisLeuAArgGluLeu	435
Db		423 CTGCAGGACCACTGGCGCTCAGGTGGCCCTGTCTGTGGGCGGAGACCTGGCGGAGCTG	482
Qy		436 ProAlaLeuValValTyGluSerGlnAaGlyGlnProValGlnAArgHisAGluHis	455
Db		483 CCGCGCTTGGTGTGTATGAGAGCCAGGCCGGGACGCGGTCCAGAGACATGAGACACAC	542
Qy		456 HisHisHisAGluHisHisHisHisTyHisHisHisHisPheTyGlnThr	470
Db		543 CACCAACATGAACATCATCACCATTTACACCACTCTACAGACA	587
RESULT 6			
ACH21474			
ID	ACH21474	standard; cDNA; 427 BP.	
XX	ACH21474;		
AC			
XX			
DT	13-OCT-2003	(first entry)	
XX			
DE	Human adult liver cDNA #1086.		
XX			
KW	Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;		
KW	genome mapping; biodiversity; genetic disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	US2003073623-A1.		
XX			
PD	17-APR-2003.		
XX			
PF	30-JUL-2001; 2001US-00918995.		
XX			
PR	30-JUL-2001; 2001US-00918995.		
XX			
PA	(DRMA/) DRMANAC R T.		
PA	(LABA/) LABAT I.		
PA	(STAC/) STACHE-CRAIN B.		
PA	(DICK/) DICKSON M C.		
PA	(JONE/) JONES L W.		
XX			
PI	Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;		
XX			
DR	WPI, 2003-615964/58.		
PT	New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene		

PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.

XX Claim 1; SEQ ID NO 8686; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623

XX Sequence 427 BP; 84 A; 177 C; 112 G; 53 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	3.87e-124	Length:	427
Score:	138.00	Matches:	138
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	29.36%	Indels:	0
DB:	9	Gaps:	0

US-09-993-966-7 (1-470) x ACH21474 (1-427)

QY 285 LeuProProArgThrSerAmpProThArGSerArGSerHISGluProGluAlaIleHis 304
DB 4 CTGCCCCCGGCACTCCATCCACATCGATCGCTCCATGAGCGGAGAACCATCCAC 63
QY 305 ILeProHISArgLySProGInGlyValASPProAlaSerPheHISpHeLeuAmpThPro 324
DB 64 ATCCCAACACCGAAGCCCAAGGCGTGACCCGCGCTCTTCCACTTCTTGACACCCCA 123
QY 325 ILaIaIyValSerGluLeuGInGlyArGLeuArGlyThrGInAPGlySerLyvHis 344
DB 124 ATGCGCAAGGCTCAAGAGCTCAAGAACGGCTCCGGGACACCGAGAGCAAGAC 183
QY 345 PheValArGSerProLySAIAGInGlyLySerValGlyValGlyHisValAlaArgGly 364
DB 184 TTGTGAGGTCCCGCAAGGCGCAAGGTGTGGGTGGGCCACGTGGCCAGAGGG 243
QY 365 AlaArgAnLySPoProLeuGlyProAlaIleProAlaValSerProSerAlaHISLeu 384
DB 244 GCAAGAAACAGCCCTCTGGGACCGGCATCCCTCGGTGTGCTCCCTCGCCACCTG 303
QY 385 AlaIaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHISLyvHis 404
DB 304 GCTCCAGCGCGGCTCTCTCCCTCTAGCCCCCTCGGGACAAAGAACCAAGAC 363
QY 405 ArgAlaLySGluSerGInGInGlyCyArGlyLeuGInAlaProLeuAlaSer 422
DB 364 CGAGCCAAAGAGACCAAGGGCTGCCGGGCTTGCAGACCACTGGCTCA 417

RESULT 7

AA592082 standard; cDNA; 1561 BP.

AA592082;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #27886.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG27895.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 1; SEQ ID NO 27886; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1561 BP; 396 A; 459 C; 397 G; 309 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.57e-93	Length:	1561
Score:	107.00	Matches:	150
Percent Similarity:	98.68%	Conservative:	0
Best Local Similarity:	98.68%	Mismatches:	1
Query Match:	22.77%	Indels:	2
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x AA592082 (1-1561)

QY 124 LeuGInCyArSPValSerMetGluGluASPserArgGInGluTrpThrPheThrLeuTyr 143

DB 3 CTCAGTGCAGCATGTTCATGAGAGACAGCCGCGAGAGTGAGACTTACCCCTTAT 62

QY 144 AspPheAspAnAnGlyLySerValThnArGlyuAspIleThrSerLeuLeuHISThrIle 163

DB 63 GACTTTACAAACAGGCAAGGTACCCGAGAGACATCACAGCTTGCTGCAACATC 122

QY 164 TyrGluValVal-AspSerSerValaenHISerProThSerSerLyvMetLeuArgVa 183

DB 123 TATGAGGTGC-GGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTGGGGT 181

Oy		183	llyvleatrrtvalalalaProAspGlySerGlnserlysvrserValleuValAamGI.nAl	203
Dd		182	AAAGCTCACCGGTGGCCCCCATATGCACGCCAGACGCAAGACGTCCTTGTAATCAGGC	241
Oy		203	aappleuginservlaarproargNlaGluThrLyfProThrgIuaAppleuarGsetr	223
Dd		242	TGACTGCAGAGGCGCAAGGCCCCCGAGCAGAAGCCAAGCCCACTGAAGCACTTGGAGCTGG	301
Oy		223	pGlulyrvylvgInArgAlAProleuarPhednGlyaspserargLeuGInGInserGI	243
Dd		302	GGAAGAAAGAACACCGAGCCCCCTCAGGTTCCAGGGGTACACGCCCTCGAGGACGTCTGG	361
Oy		243	YCyvTrYHLSHSicYvaIAspGluuAmnIllegluArgrgAmhietryLeuAAPleuAl	263
Dd		362	CtctcTACCACCATGTGGCTGATGATGAGAACITCGAGAGGAAANACCACTTAGATTCTGC	421
Oy		263	agIylllegluAentyrFthrSergInPheGlyPro	274
Dd		422	CGGGATAGAAAATACTACACGTCCTCAATTGGGCT	455
RESULT 8				
AAS78752				
ID	AAS78752	standard; cDNA; 1743 BP.		
XX				
AC	AAS78752;			
XX				
DT	13-FEB-2002	(first entry)		
DE				
XX		DNA encoding novel human diagnostic protein #14556.		
KM		Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KX		food supplement; medical imaging; diagnostic; genetic disorder; se.		
OS		Homo sapiens.		
XX				
PN	WO200175067-A2.			
PD	11-OCT-2001.			
XX				
PF	30-MAR-2001; 2001WO-US008631.			
XX				
PR	31-MAR-2000; 2000US-00540217.			
PR	23-AUG-2000; 2000US-00649167.			
XX				
PA	(HYSE-) HYSEQ INC.			
XX				
PI	Drmamac RT, Liu C, Tang YT;			
PT	WPI; 2001-639362/73.			
DR	P-PSDB; ABG14565.			
XX				
PT	New isolated polynucleotide and encoded polypeptides, useful in			
PT	diagnostics, forensics, gene mapping, identification of mutations			
PT	responsible for genetic disorders or other traits and to assess			
PT	biodiversity.			
XX				
PS	Claim 1; SEQ ID NO 14556; 103pp: English.			
XX				
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)			
CC	sequences. (I) is useful as hybridisation probes, polymerase chain			
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,			
CC	and in recombinant production of (II). The polynucleotides are also used			
CC	in diagnostics as expressed sequence tags for identifying expressed			
CC	genes. (I) is useful in gene therapy techniques to restore normal			
CC	activity of (II) or to treat disease states involving (II). (II) is			
CC	useful for generating antibodies against it, detecting or quantitating a			
CC	polypeptide in tissue, as molecular weight markers and as a food			
CC	supplement. (II) and its binding partners are useful in medical imaging			
CC	of sites expressing (II), (I) and (II) are useful for treating disorders			
CC	Involving aberrant protein expression or biological activity. The			
CC	polypeptide and polynucleotide sequences have applications in			
CC	diagnostics/forensics, gene mapping, identification of mutations			
CC				

CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS61197-AAS94564 represent novel human diagnostic
CC	coding sequences of the invention. Note: The sequence data for this
CC	patent did not appear in the printed specification, but was obtained in
CC	electronic format directly from WPO at
CC	ftp.wipo.int/pub/published_pct_sequences
xx	
6Q	Sequence 1743 BP, 367 A, 562 C, 511 G, 303 T; 0 U; 0 Other;
	Alignment Scores:
	Pred. No.: 1,566-59 Length: 1743
	Score: 72.00 Matches: 72
	Percent Similarity: 100.00% Conservative: 0
	Best Local Similarity: 100.00% Mismatch: 0
	Query Match: 15.32% Indels: 0
	DB: 5 Gaps: 0
US	09-993-966-7 (1-470) x AAS78752 (1-1743)
Oy	276 SerProSerValAlaGlnIysSerGluLeuProProArgThrSerAsnProThrArgSer 295
Db	334 TCCCTTCGGTGAGCCCAAGATCAAGATGCCCCCGACCTCCATCCCATCGATCT 393
Oy	296 ArgSerHisGluProGlnAlaIleHisIleProHisArgIysProGlnIylValAspPro 315
Db	394 CGCTCCCATGAGCCGGAAGCCATCCACATCCACACCGAAAGCCCAAGCCGTGGACCG 453
Oy	316 AlaSerPheHisPheLeuAspThrProIleAlaIysValSerGluLeuGlnIlnArgLeu 335
Db	454 GCCTCTTCCACTTCTCTTGACACCCCATCGCCAAAGTCTCAGAGCTCCAGACACGGCTC 513
Oy	336 ArgGlyThrGlnAspGlySerIysHisPheValArg 347
Db	514 CGGGGCAACCAAGACGAGGACAGCACTTGTGAGG 549
RESULT 9	
ID	AAS92081 standard; cDNA, 1743 BP.
xx	AAS92081
xx	AAS92081;
xx	13-FEB-2002 (first entry)
DT	
xx	DNA encoding novel human diagnostic protein #27885.
DE	
xx	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
xx	
OS	Homo sapiens.
xx	
xx	WO200175067-A2.
xx	
PD	11-OCT-2001.
xx	
xx	30-MAR-2001; 2001WO-US008631.
xx	
PR	31-MAR-2000; 2000US-00540217.
xx	
PR	23-AUG-2000; 2000US-00649167.
xx	
PA	(HYSE-) HYSEQ INC.
xx	
xx	Dmanac RT, Liu C, Tang YT;
xx	
DR	WPI: 2001-639362/73.
xx	
xx	P-PSDB; AEG27894.
xx	
xx	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
xx	
xx	responsible for genetic disorders or other traits and to assess
xx	biodiversity.
xx	
xx	Claim 1; SEQ ID NO 27885; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 1.56e-59 Length: 1743
Score: 72.00 Matches: 72
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.32% Indels: 0
Gaps: 0

US-09-993-966-7 (1-470) x AAS92081 (1-1743)

QY 276 SerProSerValAlaGlnIlySerGluLeuProProArgThrSerAsnProThrArgSer 295
Db 334 TCCCTTCCTCCGCGCCGAGAGTCAGAACTGCCCCCGACCTCCAAATCCACATCGATCT 393
QY 296 ArgSerHisGluProGluAlaIleHisIleProHisArgIlyProGlnIlyValAspPro 315
Db 394 CGCTCCCATAGACCGGAGAGCCATCCATCCACACGAAAGCCCAAGGGGTGAGCCG 453
QY 316 AlaSerPheHisPheLeuAspThrProIleAlaIlyValSerGluLeuGlnArgLeu 335
Db 454 GCCTCTTCACCTTCCTTGACACCCCAATCCCAAGTCTCAGAGCTCCAGCAACGGCTC 513
QY 336 ArgGlyThrGlnArgIlySerIlyHisPheValArg 347
Db 514 CGGGGACCCAGAGCGGAGAGCAACACTTGTGAG 549
RESULT 10
AD31966
ID AD31966 standard; cDNA; 1743 BP.
XX
AC AD31966;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel cDNA contig sequence, SEQ ID NO:2048.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnery;
KW antidiabetic; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN MO2003029271-A2.

XX 10-Apr-2003.
PD
XX
XX 24-SEP-2002; 2002MO-US030474.
PF
XX 24-SEP-2001; 2001US-0324631P.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI: 2003-371981/35.
DR P-PSDB; ADC32733.
XX
XX New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
XX Example 2; SEQ ID NO 2048; 1185pp; English.
PS
XX
XX The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.56e-59 Length: 1743
Score: 72.00 Matches: 72
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.32% Indels: 0
Gaps: 0

US-09-993-966-7 (1-470) x AD31966 (1-1743)

QY 276 SerProSerValAlaGlnIlySerGluLeuProProArgThrSerAsnProThrArgSer 295
Db 334 TCCCTTCCTCCGCGCCGAGAGTCAGAACTGCCCCCGACCTCCAAATCCACATCGATCT 393
QY 296 ArgSerHisGluProGluAlaIleHisIleProHisArgIlyProGlnIlyValAspPro 315

```
DB 394 CGCTCCATGAGCCGGAAGCCATCCACATCCGACCGAAGCCGAGGAGCCG 453
QY 316 AlaSerPheHisPheLeuAspThrProIleAlaIysValSerGluLeuGlnIleArgLeu 335
DB 454 GCGTCCCTTCCACTTCCTTGAACCCCAATGCCAAGGCTCAGAGCTCCAGCAAGGCTC 513
QY 336 ArgGlyThrGlnIleAspGlySerLysHisPheValArg 347
DB 514 CCGGGCACCCAGAGCGGAGCAGACACTTGTGTGAG 549

RESULT 11
AAL51522
ID AAL51522 standard; DNA, 148 BP.
XX
AC AAL51522;
XX
DT 24-APR-2003 (first entry)
XX
DE Human Nkd (hNkd) gene exon 8 (coding region).
XX
KW Human; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;
KW colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
KW promoter.
XX
OS Homo sapiens.
XX
PN WO200292832-A2.
XX
PD 21-NOV-2002.
XX
PF 27-NOV-2001; 2001WO-US044092.
XX
PR 27-NOV-2000; 2000US-0252884P.
PR 16-MAY-2001; 2001US-0291109P.
PR 01-OCT-2001; 2001US-0325571P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Rohan M, Chan V, Yan D;
XX
DR WPI; 2003-129303/12.
XX
PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
PT useful for treating cancer involving aberrant Wnt signalling, e.g. colon
PT cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
PS Example 3; Fig 7; 99pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the human
CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence represents exon 8
CC (coding region) from the human Nkd gene
XX
SQ Sequence 148 BP; 35 A; 48 C; 38 G; 27 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,06e-38 Length: 148
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
DB: 10 Gaps: 0

US-09-993-966-7 (1-470) x AAL51522 (1-148)

QY 155 AspIleThrSerLeuLeuHisThrIleTyrGluValAlaSerSerValAsnHisSer 174
DB 1 GACATCACCAAGCTTGACACACCATGTATGAGGTGAGCTCTGTCAACCACTCC 60
QY 175 ProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySerGlnSer 194
```

```
DB 61 CCAACATCCAGCAAGATGCTGGGTTAAAGCTCACCGTGGCCCCGATGGACGACAGC 120
QY 195 LysArgSerValLeuValAsnGlnAla 203
DB 121 AAGAGAGAGCGTCCCTTGTCAATCAGGCT 147

RESULT 12
AAL51512
ID AAL51512 standard; DNA, 1401 BP.
XX
AC AAL51512;
XX
DT 24-APR-2003 (first entry)
XX
DE Mouse Nkd protein gene sequence.
XX
KW Mouse; gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;
KW colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
KW murine.
XX
OS Mus sp.
XX
PN WO200292832-A2.
XX
PD 21-NOV-2002.
XX
PF 27-NOV-2001; 2001WO-US044092.
XX
PR 27-NOV-2000; 2000US-0252884P.
PR 16-MAY-2001; 2001US-0291109P.
PR 01-OCT-2001; 2001US-0325571P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Rohan M, Chan V, Yan D;
XX
DR WPI; 2003-129303/12.
XX
PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
PT useful for treating cancer involving aberrant Wnt signalling, e.g. colon
PT cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
PS Example 2; Fig 4; 99pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the human
CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence represents the mouse
CC Nkd gene
XX
SQ Sequence 1401 BP; 336 A; 430 C; 416 G; 219 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.6e-37 Length: 1401
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
DB: 10 Gaps: 0

US-09-993-966-7 (1-470) x AAL51512 (1-1401)

QY 131 GluGluAspSerArgGlnGluTyrTrpThrLeuTyrAspPheAspAsnAsnGlyLys 150
DB 391 GAGGAGGACAGCCGGGACAGTGACTTCACTCTATATACCTTCGACACATGGGAAA 450
QY 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValAlaSerSerSer 170
DB 451 GTGACCCGCTAGGACATTACCACTTGCTGCATACATCATGAGAGGTTGACTCTCT 510
QY 171 ValAsnHisSerProThrSerSerLys 179
```

```
Db 511 GTGACCATTCCTCCCACTCAGCAAG 537
|||||
RESULT 13
AAH28343
ID AAH28343 standard; DNA; 1416 BP.
XX
AC AAH28343;
XX
DT 05-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a murine mnkd protein.
XX
KM Dishevelled-associated protein 1A; DAPIA; mnkd; dishevelled protein;
KW Wnt signalling; colon cancer; cancer; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1416
FT /cag= a
FT /product= "mnkd"
XX
MO200144279-A2.
XX
PD 21-JUN-2001.
XX
PE 05-DEC-2000; 2000MO-US032986.
XX
PR 17-DEC-1999; 99US-0172434P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Yan D, Williams LT;
XX
DR WPI; 2001-425440/45.
XX
DR P-PSDB; AAB84645.
XX
PT Novel mammalian dishevelled-associated proteins, mnkd and DAPIA, useful
PT for inhibiting Wnt signalling in mammalian cells, and thus for treating
PT colon cancer.
XX
PS Claim 2; Fig 1; 67pp; English.
XX
CC The present sequence encodes a murine mnkd protein. The specification
CC describes mnkd and dishevelled-associated protein 1A (DAPIA) proteins,
CC that interact with mammalian dishevelled protein. mnkd is useful for
CC inhibiting Wnt signalling in mammalian cells, and thus for treating colon
CC cancer. mnkd is also useful for activating the JNK pathway. mnkd and
CC DAPIA are also useful for screening drugs that are useful for treating
CC cancer.
XX
SQ Sequence 1416 BP; 339 A; 434 C; 419 G; 224 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.62e-37 Length: 1416
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
Gaps: 0
DB: 4

US-09-993-966-7 (1-470) x AAH28343 (1-1416)

QY 131 GUGUuApSerArGInGInTPrThrPheThrLeuYrAspPheAspAsnAsnGlyLys 150
|||
Db 391 GAGGAGACACCGCGCAAGATGACCTTCACTCATATGACTTCGACCAACAATGGGAAA 450
|||
QY 151 ValThrArgGluuApIleThrSerIleuHisThrIleYrGluValValAspSerSer 170
|||
Db 451 GTGACCGGTGAGACATTTACAGCTTGCTGCATACCATTTGAAGTGGTGAAGTGGTCCCTCT 510
|||
QY 171 ValAsnHisSerProThrSerSerLys 179
|||
```

```
Db 511 GTGACCATTCCTCCCACTCAGCAAG 537
|||||
RESULT 14
ABQ73820
ID ABQ73820 standard; cDNA; 1950 BP.
XX
AC ABQ73820;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human colon specific nucleic acid (CSNA) SEQ ID NO:126.
XX
KM Human; colon specific nucleic acid; colon specific polypeptide; CSP;
KW CSNA; colon specific gene; CSG; colon cancer; gene therapy; vaccine;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1950
FT /cag= a
FT /product= "CSNA"
XX
MO200248370-A2.
XX
PD 20-JUN-2002.
XX
PE 30-OCT-2001; 2001MO-US051341.
XX
PR 31-OCT-2000; 2000US-0244717P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Ghosh MG, Liu C;
XX
DR WPI; 2002-583520/62.
XX
PT Colon specific polypeptides and polynucleotides useful for detecting,
PT diagnosing, monitoring, treating, staging and predicting cancers in
PT humans having cancer and non-cancerous colon disease.
XX
PS Claim 1; Page 199; 243pp; English.
XX
CC ABQ73695 to ABQ73841 represent human colon specific nucleic acid (CSNA)
CC sequences, and ABP51826 to ABP51928 represent human colon specific
CC polypeptide (CSP) sequences from the present invention. CSNA and CSP
CC sequences have cytostatic activity, and can be used in gene therapy,
CC antisense therapy and in vaccines. CSNA and CSP sequences can be used for
CC diagnosing and monitoring the presence and metastases of colon cancer in
CC a patient, by determining an amount of CSP or CSNA in a sample of a
CC patient, and comparing it to the amount of colon specific marker in a
CC normal control, where a difference in the amount of the nucleic acid or
CC the polypeptide in the sample compared to that of normal control is
CC associated with presence of colon cancer. CSP and CSNA sequences can be
CC used for producing engineered colon tissue for treatment and research.
CC CSNA sequences are useful for producing transgenic animals and cells and
CC also in gene therapy
XX
SQ Sequence 1950 BP; 459 A; 494 C; 530 G; 467 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.89e-34 Length: 1950
Score: 46.00 Matches: 46
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.79% Indels: 0
Gaps: 0
DB: 6

US-09-993-966-7 (1-470) x ABQ73820 (1-1950)

QY 20 G1YAspSerPheAlaValISerAlaATrPA1ArgHyGlyIleGluGluTrpIleGly 39
|||
Db 838 GGTGACAGCTTTCGCGGAGCGCTGCTGGGCTCGAAGGCGATCGAGAGTGAATCGG 897
|||
QY 40 ArgGlnArgCysProGlyGlyValISerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
|||
Db 898 AGACAGCGCTGCCGCGGCGGTCTCGGACCCGACAGCTGCGGCTTGGCGGACCATATA 957
|||
```


QY 60 GlyArgSerThrArgIle 65
 DB 958 GGCGAAGCACCCGGGAA 975

RESULT 15
 ID ACH74795 standard; DNA, 598 BP.
 XX ACH74795;
 AC ACH74795;
 XX 29-JUL-2004 (first entry)
 DT 29-JUL-2004 (first entry)
 XX Human genome derived single exon probe #7990.
 DE Human; probe; ss; gene expression; single exon probe; microarray;
 KM alternative splicing event; genomic alteration.
 XX Homo sapiens.
 OS Homo sapiens.
 XX US2003194704-A1.
 XX 16-OCT-2003.
 PD 03-APR-2002; 2002US-00029386.
 PF 03-APR-2002; 2002US-00029386.
 PR 03-APR-2002; 2002US-00029386.
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX Penn SG, Rank DR, Hanzel DK;
 PI WPI; 2004-119264/12.
 DR New human genome-derived single exon nucleic acid probes useful for human
 XX gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 PS Claim 15; SEQ ID NO 7990; 80bp; English.
 XX The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above, a
 CC method of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,

CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20030194704
 XX

SQ Sequence 598 BP, 84 A; 232 C; 162 G; 120 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.77e-34 Length: 598
 Score: 45.00 Matches: 45
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.57% Indels: 0
 DB: 12 Gaps: 0

US-09-993-966-7 (1-470) x ACH74795 (1-598)

QY 20 GlyArgSerPheAlaValSerAlaIleTPAlaArgIleGlyIleGluIleTPIleGly 39
 DB 229 GGTGACAGCTTCGCGGAGCGCTGCTGGGCTCGGAGGCGCATCGAGAGTGATCGGG 288

QY 40 ArgGlnArgCysProGlyIleValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
 DB 289 AGACAGCGCTGCCCGGCGGTGTCTCGGAGACCCCGACAGCTTGGCGGCGACCATTA 348

QY 60 GlyArgSerThrArg 64
 DB 349 GGCGAAGCACCCGG 363

Search completed: December 30, 2004, 09:47:45
 Job time: 860 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 09:00:11 ; Search time 142 Seconds
(without alignments)
2352.611 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 470

Sequence: 1 MGLKSKPAVCKRRSPG.....RHHHHHHHHHHHHFYQT 470

Scoring table: (OLIGO)
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 35539441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-O/cgn2.1/USPTO.spool.p/US0993966/runat.29122004.130830.9412/app.query.fasta_1.647
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=oligo.rnt -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=coligo -TRANS=human40.cai
-LIST=45 -DOCLIN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0993966 @CEN 1.1 128 @runat.29122004.130830.9412 -NCPU=6 -ICPU=3
-NO MAMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383	81.5	1438	4	US-09-506-066E-5
2	45	9.6	1731	4	US-09-506-066E-3
3	17	3.6	1285	4	US-09-506-066E-7
4	17	3.6	1307	4	US-09-506-066E-9
5	8	1.7	75	2	US-08-447-169A-125
6	8	1.7	75	2	US-08-233-012C-125
7	8	1.7	75	4	US-09-860-474-125
8	8	1.7	428	4	US-09-641-638-497
9	8	1.7	428	4	US-10-170-097-497
10	8	1.7	449	4	US-09-976-594-405
11	8	1.7	455	4	US-09-621-976-1583
12	8	1.7	458	4	US-09-270-767-3943

13	8	1.7	458	4	US-09-270-767-19325	Sequence 19225, A
14	8	1.7	462	4	US-09-270-767-5688	Sequence 5698, Ap
15	8	1.7	482	4	US-09-270-767-20580	Sequence 20980, A
16	8	1.7	501	4	US-09-583-110-1646	Sequence 1646, Ap
17	8	1.7	531	4	US-09-621-976-586	Sequence 586, App
18	8	1.7	533	4	US-09-270-767-27303	Sequence 27303, A
19	8	1.7	639	4	US-09-270-767-14169	Sequence 14169, A
20	8	1.7	683	4	US-09-270-767-11684	Sequence 11684, A
21	8	1.7	732	4	US-09-248-796A-1692	Sequence 1692, Ap
22	8	1.7	924	4	US-09-270-767-87	Sequence 87, Appl
23	8	1.7	924	4	US-09-270-767-15369	Sequence 15369, A
24	8	1.7	1155	3	US-09-439-313-373	Sequence 373, App
25	8	1.7	1155	3	US-09-352-616A-373	Sequence 373, App
26	8	1.7	1155	4	US-09-289-198-301	Sequence 301, App
27	8	1.7	1155	4	US-09-636-215-373	Sequence 373, App
28	8	1.7	1155	4	US-09-685-166A-373	Sequence 373, App
29	8	1.7	1155	4	US-09-429-755-301	Sequence 301, App
30	8	1.7	1155	4	US-09-679-426-373	Sequence 373, App
31	8	1.7	1225	4	US-09-270-767-10546	Sequence 10546, A
32	8	1.7	1366	4	US-09-149-476-49	Sequence 49, Appl
33	8	1.7	1512	3	US-09-439-313-368	Sequence 368, App
34	8	1.7	1512	3	US-09-062-451-294	Sequence 294, App
35	8	1.7	1512	3	US-09-352-616A-368	Sequence 368, App
36	8	1.7	1512	4	US-09-289-198-294	Sequence 294, App
37	8	1.7	1512	4	US-09-636-215-368	Sequence 368, App
38	8	1.7	1512	4	US-09-685-166A-368	Sequence 368, App
39	8	1.7	1512	4	US-09-429-755-294	Sequence 294, App
40	8	1.7	1512	4	US-09-679-426-368	Sequence 368, App
41	8	1.7	1764	4	US-09-248-796A-4659	Sequence 4659, Ap
42	8	1.7	1787	2	US-08-808-982-2	Sequence 2, Appl
43	8	1.7	1787	3	US-09-306-902A-2	Sequence 2, Appl
44	8	1.7	1853	3	US-09-439-313-369	Sequence 369, App
45	8	1.7	1853	3	US-09-062-451-295	Sequence 295, App

ALIGNMENTS

RESULT 1
US-09-506-066E-5
; Sequence 5, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506, 066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120, 646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8) ..(1418)
; OTHER INFORMATION: Nkd1 coding sequence
US-09-506-066E-5

Alignment Scores:
Pred. No.: 0
Score: 383.00
Percent Similarity: 99.58%
Best Local Similarity: 99.58%
Query Match: 81.49%
DB: 4
US-09-993-966-7 (1-470) x US-09-506-066E-5 (1-1438)

Length: 1438
Matches: 469
Conservative: 0
Mismatch: 1
Indels: 2
Gaps: 0

```

QY 1 MetGlyLysLeuHisSerIysProAlaAlaValCysLysArgArgIuserProGluGly 20
Db 8 ATGGGGAAATCTTCACTCAAGCCGCGCGCTGTGAGACGACAGAGAGCCCGAAGGT 67
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluGluTrpIleGlyArg 40
Db 68 GACAGCTTCGCGCGTGAAGGCTGCTGGGCTCGAAGGCGCATCGAGGTGATCGGAGGA 127
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db 128 CAGGCTGCCCGCGCGGTGTCTCGGAGCCCGACAGCTGCGGTGGCGGACCATGAGC 187
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCGGAGGCTGTGGCGAGCTGTGAGAGACAGCTCAAGCAGAGAGAGAG 247
QY 81 AspArgPheArgLeuGluValAlaLeuProGluTrpThrAspGlyLeuGlySerGly 100
Db 248 GACGACTTTCGCTGGAAGTGGCCCTGCTCTGAGAAAGCTGACGGGCTGGCGAGCGA 307
QY 101 AspGluLysLysMetGluArgValSerGluProCysProGlySerLysLysGluLeuLys 120
Db 308 GATGAGAGAGAGATGAGAGAGTGAAGGACCTGCGCGAGCTCCAAAGACAGCTTAAG 367
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluLeuAspSerArgGlnGluTrpThrPhe 140
Db 368 TTTTAAAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCCGCAGAGTGAACCTTC 427
QY 141 ThrLeuTrpAspPheAspAsnAsnGlyLysValThrArgLysAspIleThrSerLeuLeu 160
Db 428 ACCCTGATGACTTTGACAACAAAGGTCACCCGAGAGGACATCACAGCTTGCTG 487
QY 161 HisThrIleTrpGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 488 CACACCATCATGAGGTGTGAGCTCTGTCTCAACCACTCCCAACATCCAGCAAGATG 547
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200
Db 548 CTGGGGGTAAAGCTCACCGTGTGCCCCGATGTGCGAGCCAGCAAGAGAGAGAGCTTGTG 607
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgLysGluThrLysProThrGluAspLeu 220
Db 608 AATCAGGCTGACTGACAGAGCGCAAGGCCCGGACAGAGACCAAGCCCATGAGAGACTG 667
QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGGAGCTGGAGAAAGAGAGCGAGCGCCGCTCAGGTTCAGGGGTGACAGCGCGCTGAG 727
QY 241 GluSerGlyCysTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisGlyPhe 260
Db 728 CAGTCTGGCTGCTCAACCATGTGCTGATGAGAAACATCGAGAGAGAAACCATCTTA 787
QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db 788 GATCTCCCGGAGATAGAAACTACAGCTCCCAATTTGGGCTCGCTCCCTTCGTGGGCC 847
QY 281 GluLysSerGluLeuProProArgTrpThrSerAspProThrArgSerArgHisGluPro 300
Db 848 CAGAGTCAAGAACTGCCCCCGCACCTT-CATCTCCACTGATCTGCTCCCATGAGCC 906
QY 300 OGAlaAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db 907 GGAAGCATTCACATCCACACCGAAAGCCCAAGGGGTGAGCCCGGCTTCCTTCCACTT 966
QY 320 eLeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAs 340
Db 967 CCTTGACACCCCAATCGCAAGTCTCAAGACTCCACAGCAAGGCTCGGGGCACTCAGGA 1026
QY 340 PGIySerLysHisPheValArgSerProLysValGlnGlyLysSerValGlyValHis 360
Db 1027 CGGAGAGCAAGACTTTGTGAGTCCCCCAAGGCCAGAGGCAAGGTGTGGTGGGCCCA 1086
QY 360 sValAlaArgGlyValaArgLeuLysProProLeuGlyProAlaIleProAlaValSerPr 380

```

```

Db 1087 CTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGCCCGGCATCTCGGTGTCTCCC 1146
QY 380 oSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1147 CTCGCCCACTGTGGTCCAGCCCGGCCCTCTCCCTCCCTAGCCCTCCGGGCAAA 1206
QY 400 sLysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420
Db 1207 GAAGCACAAGACCGAGCCAGAGAGAGAGAGAGAGGCTGCGGGGCTTGAGGCCACT 1266
QY 420 uAlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuValVa 440
Db 1267 GGCCTCAGGTGGCGCTGTCTGTGGGGGAGAGACACTGCGGAGACTGCGGCTTGGTGT 1326
QY 440 lTyrgLysGlnAlaGlyGlnProValGlnArgHisGluHisHisHisGluHis 460
Db 1327 GTATGAGAGCCAGGCGGGGAGCGGGTCCAGAGACATGAGCACACCAACCATGACA 1386
QY 460 sHisHisIleTrpHisHisPheTrpGluThr 470
Db 1387 TCACCACTATTACCACTTCTTACAGACA 1417

RESULT 2
US-09-506-066E-3
: Sequence 3, Application US/09506066E
: Patent No. 6630323
: GENERAL INFORMATION:
: APPLICANT: Scott, Matthew
: APPLICANT: Wharton, Keith
: APPLICANT: Zeng, Wenlin
: TITLE OF INVENTION: Naked Cuticle Genes and their Uses
: FILE REFERENCE: STAN-121
: CURRENT APPLICATION NUMBER: US/09/506, 066E
: PRIOR FILING DATE: 2000-02-17
: PRIOR APPLICATION NUMBER: 60/120, 646
: NUMBER OF SEQ. ID NOS: 15
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 1731
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (140)...(1553)
: OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-3

Alignment Scores:
Pred. No.: 2,28E-33 Length: 1731
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.57% Indels: 0
Gaps: 0

US-09-993-966-7 (1-470) x US-09-506-066E-3 (1-1731)
QY 1 MetGlyLysLeuHisSerIysProAlaAlaValCysLysArgArgIuserProGluGly 20
Db 140 ATGGGAAATCTTCACTCAAGCCGCGCGCTGTGAGACGACAGAGAGCCCGAAGGT 199
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluGluTrpIleGlyArg 40
Db 200 GACAGCTTCGCGCGTGAAGGCTGCTGGGCTCGAAGGCGCATCGAGGTGATCGGAGG 259
QY 41 GluArgCysProGly 45
Db 260 CAGCGCTGTCAAGC 274

RESULT 3
US-09-506-066E-7

```

```

; Sequence 7, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STRAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1285
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1282)
; OTHER INFORMATION: Nkd2 coding sequence
US-09-506-066E-7

```

```

Alignment Scores:
Pred. No.: 9.31e-07 Length: 1285
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.62% Indels: 0
DB: 4 Gaps: 0

```

US-09-993-966-7 (1-470) x US-09-506-066E-7 (1-1285)

```

Qy 254 GUAARGAghshstYrLeuApLeuA1gY11egLubantYrTrser 270
Db 644 GAGCGGAAACCACTACCTAGACCTTGCTGGCATCGAGACTACACATCT 694

```

```

RESULT 4
US-09-506-066E-9
; Sequence 9, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STRAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-506-066E-9

```

```

Alignment Scores:
Pred. No.: 9.46e-07 Length: 1307
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.62% Indels: 0
DB: 4 Gaps: 0

```

US-09-993-966-7 (1-470) x US-09-506-066E-9 (1-1307)

```

Qy 254 GUAARGAghshstYrLeuApLeuA1gY11egLubantYrTrser 270
Db 364 GAGCGGAAACCACTACCTAGACCTTGCTGGCATCGAGACTACACATCT 414

```

```

RESULT 5
US-08-447-169A-125/c
; Sequence 125, Application US/08447169A
; Patent No. 581533
; GENERAL INFORMATION:
; APPLICANT: JANIC, N. and GOLD, L.
; TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE
; TITLE OF INVENTION: LIGANDS TO VASCULAR ENDOTHELIAL
; NUMBER OF INVENTION: GROWTH FACTOR (VEGF)
; NUMBER OF SEQUENCES: 242
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Place, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,169A
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/233,012
; FILING DATE: 25-APRIL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/205,515
; FILING DATE: 03-MARCH-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/964,624
; FILING DATE: 21-OCTOBER-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 125:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: All C's are 2'-NH2 cytosine
; FEATURE:
; OTHER INFORMATION: All U's are 2'-NH2 uracil
US-08-447-169A-125

```

```

Alignment Scores:
Pred. No.: 24.6 Length: 75
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70% Indels: 0
DB: 1 Gaps: 0

```

US-09-993-966-7 (1-470) x US-08-447-169A-125 (1-75)

```

Qy 387 SerProAlaLeuEupProSerLeu 394

```

Db 43 TCACCTGCGCTCTTCCTCTTG 20

RESULT 6
US-08-233-012C-125/c

; Sequence 125, Application US/08233012C
; Patent No. 5849479

; GENERAL INFORMATION:

; APPLICANT: JANJIC, N. and GOLD, L.

; TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE

; TITLE OF INVENTION: LIGANDS TO VASCULAR

; TITLE OF INVENTION: ENDOTHELIAL GROWTH FACTOR

; NUMBER OF SEQUENCES: 146

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Swanson & Bratschun, L.L.C.

; STREET: 8400 E. Prentice Place, Suite 200

; CITY: Englewood

; STATE: Colorado

; COUNTRY: USA

; ZIP: 80111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG

; MEDIUM TYPE: storage

; COMPUTER: IBM compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: WordPerfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233.012C

; FILING DATE: 25-APRIL-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/714,131

; FILING DATE: 10-JUNE-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/536,428

; FILING DATE: 11-JUNE-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/964,624

; FILING DATE: 21-OCTOBER-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Barry J. Swanson

; REGISTRATION NUMBER: 33,215

; REFERENCE/DOCKET NUMBER: NEX14

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 793-3333

; TELEFAX: (303) 793-3433

; INFORMATION FOR SEQ ID NO: 125:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 75 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; OTHER INFORMATION: All C's are 2'-NH2 cytosine

; FEATURE:

; OTHER INFORMATION: All U's are 2'-NH2 uracil

; US-08-233-012C-125

; Alignment Scores:

; Pred. No.: 24.6

; Score: 8.00

; Percent Similarity: 100.00%

; Best Local Similarity: 100.00%

; Query Match: 1.70%

; DB: 2

; US-09-993-966-7 (1-470) x US-08-233-012C-125 (1-75)

Qy 387 SetProAlaLeuProSerLeu 394

Db 43 TCACCTGCGCTCTTCCTCTTG 20

US-09-860-474-125/c

; Sequence 125, Application US/09860474

; Patent No. 6696252

; GENERAL INFORMATION:

; APPLICANT: GOLD, L. and JANJIC, N.

; TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE LIGANDS

; TITLE OF INVENTION: TO VASCULAR ENDOTHELIAL GROWTH

; FACTOR (VEGF)

; NUMBER OF SEQUENCES: 242

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Swanson & Bratschun, L.L.C.

; STREET: 1745 Shea Center Drive, Suite 330

; CITY: Highlands Ranch

; STATE: Colorado

; COUNTRY: USA

; ZIP: 80128

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage

; COMPUTER: IBM compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: WordPerfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/860,474

; FILING DATE: 18-May-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/156,824

; FILING DATE: 18-SEPTEMBER-1998

; APPLICATION NUMBER: 08/447,169

; FILING DATE: 19-MAY-1995

; APPLICATION NUMBER: 08/233,012

; FILING DATE: 25-APRIL-1994

; APPLICATION NUMBER: 08/205,515

; FILING DATE: 03-MARCH-1994

; APPLICATION NUMBER: 07/964,624

; FILING DATE: 21-OCTOBER-1992

; APPLICATION NUMBER: 07/714,131

; FILING DATE: 10-JUNE-1991

; APPLICATION NUMBER: 07/536,428

; FILING DATE: 11-JUNE-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Barry J. Swanson

; REGISTRATION NUMBER: 33,215

; REFERENCE/DOCKET NUMBER: NEX14/CIP-CON

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 793-3333

; TELEFAX: (303) 793-3433

; INFORMATION FOR SEQ ID NO: 125:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 75 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; OTHER INFORMATION: All C's are 2'-NH2 cytosine

; FEATURE:

; OTHER INFORMATION: All U's are 2'-NH2 uracil

; US-09-860-474-125

; Alignment Scores:

; Pred. No.: 24.6

; Score: 8.00

; Percent Similarity: 100.00%

; Best Local Similarity: 100.00%

; Query Match: 1.70%

; DB: 4

; US-09-993-966-7 (1-470) x US-09-860-474-125 (1-75)

Qy 387 SetProAlaLeuProSerLeu 394

Db 43 TCACCTGCGCTCTTCCTCTTG 20

```
RESULT 8
US-09-641-638-497
; Sequence 497, Application US/09641638
; Patent NO. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.05ICP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 497
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 281
; OTHER INFORMATION: 10-15-281 : polymorphic base G or T
; NAME/KEY: misc_binding
; LOCATION: 261..280
; OTHER INFORMATION: 10-15-281.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 282..301
; OTHER INFORMATION: 10-15-281.misl, potential
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 409..428
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 269..293
; OTHER INFORMATION: 10-15-281 potential probe
US-09-641-638-497

Alignment Scores:
Pred. No.: 128 Length: 428
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70% Indels: 0
DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-641-638-497 (1-428)
Qy 388 ProAlaLeuLeuProSerLeuAla 395
Db 89 CCAGCCTGCTCTCTCTCTGACC 112

RESULT 9
US-10-170-097-497
; Sequence 497, Application US/10170097
; Patent NO. 6794143
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
```

```
;; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;; FILE REFERENCE: GEN-T114XC2D1
;; CURRENT APPLICATION NUMBER: US/10/170,097
;; CURRENT FILING DATE: 2002-06-10
;; PRIOR APPLICATION NUMBER: US 09/641,638
;; PRIOR FILING DATE: 2000-08-16
;; PRIOR APPLICATION NUMBER: US 09/502,330
;; PRIOR FILING DATE: 2000-02-11
;; PRIOR APPLICATION NUMBER: US 60/133,200
;; PRIOR FILING DATE: 1999-05-07
;; PRIOR APPLICATION NUMBER: US 09/275,267
;; PRIOR FILING DATE: 1999-03-23
;; PRIOR APPLICATION NUMBER: US 60/119,917
;; PRIOR FILING DATE: 1999-02-12
;; NUMBER OF SEQ ID NOS: 1304
;; SOFTWARE: Patent.pm
;; SEQ ID NO 497
;; LENGTH: 428
;; TYPE: DNA
;; ORGANISM: Homo Sapiens
;; FEATURE:
;; NAME/KEY: allele
;; LOCATION: 281
;; OTHER INFORMATION: 10-15-281 : polymorphic base G or T
;; FEATURE:
;; NAME/KEY: misc_binding
;; LOCATION: 261..280
;; OTHER INFORMATION: 10-15-281.misl, potential
;; FEATURE:
;; NAME/KEY: misc_binding
;; LOCATION: 282..301
;; OTHER INFORMATION: 10-15-281.misl, potential
;; NAME/KEY: primer_bind
;; LOCATION: 1..20
;; OTHER INFORMATION: upstream amplification primer
;; NAME/KEY: primer_bind
;; LOCATION: 409..428
;; OTHER INFORMATION: downstream amplification primer, complement
;; NAME/KEY: misc_binding
;; LOCATION: 269..293
;; OTHER INFORMATION: 10-15-281 potential probe
US-10-170-097-497

Alignment Scores:
Pred. No.: 128 Length: 428
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70% Indels: 0
DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-10-170-097-497 (1-428)
Qy 388 ProAlaLeuLeuProSerLeuAla 395
Db 89 CCAGCCTGCTCTCTCTCTGACC 112

RESULT 10
US-09-976-594-405
; Sequence 405, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
```

```
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 405
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 199286.3
; LOCATION: 397
; NAME/KEY: unsure
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-405

Alignment Scores:
Pred. No.: 134          Length: 449
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-976-594-405 (1-449)

QY 384 LeuAlaAlaSerProAlaLeuLeu 391
Db 121 CTGGCCGGAAGCCCCGCGTGTCTA 144

RESULT 11
US-09-621-976-1583/c
; Sequence 1583, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1583
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 143..433
; NAME/KEY: sig_peptide
; LOCATION: 143..217
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 6
; OTHER INFORMATION: seq LIGLIVAVATVHL/VI
US-09-621-976-1583

Alignment Scores:
Pred. No.: 136          Length: 455
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-621-976-1583 (1-455)

QY 334 ArgLeuArgGlyThrGlnAspGly 341
Db 332 CGACTCCGGGGAACTCAAGATGCT 309

RESULT 12
US-09-270-767-3943

; Sequence 3943, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3943
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-3943

Alignment Scores:
Pred. No.: 137          Length: 458
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-270-767-3943 (1-458)

QY 456 HisHisHisGluHisHisHis 463
Db 300 CATCACACGAGCACACACACAC 323

RESULT 13
US-09-270-767-19225
; Sequence 19225, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 19225
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-19225

Alignment Scores:
Pred. No.: 137          Length: 458
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-270-767-19225 (1-458)

QY 456 HisHisHisGluHisHisHis 463
Db 300 CATCACACGAGCACACACACAC 323

RESULT 14
US-09-270-767-5698/c
; Sequence 5698, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
```



```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5698
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-5698

```

```

Alignment Scores:
Pred. No.:      144      Length:      482
Score:          8.00     Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     1.70%   Indels:      0
DB:              4      Gaps:        0

```

```

US-09-993-966-7 (1-470) x US-09-270-767-5698 (1-482)

```

```

QY      70 ValLeuArgAspThrLeuSergLu 77
      |||||
DB      48 GTTTAAGAGACACCTTAACGAA 25

```

RESULT 15

```

US-09-270-767-20980/c
; Sequence 20980, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 20980
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-20980

```

```

Alignment Scores:
Pred. No.:      144      Length:      482
Score:          8.00     Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     1.70%   Indels:      0
DB:              4      Gaps:        0

```

```

US-09-993-966-7 (1-470) x US-09-270-767-20980 (1-482)

```

```

QY      70 ValLeuArgAspThrLeuSergLu 77
      |||||
DB      48 GTTTAAGAGACACCTTAACGAA 25

```

```

Search completed: December 30, 2004, 13:44:46
Job time : 148 secs

```

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 07:54:31 ; Search time 5643 Seconds
(without alignments)
3035.030 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 470
Sequence: 1 MGKLHKKPAVCKRRSPG.....RHHHHHHHHHHYHFYQT 470

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xip
-Q/cgnt 1/USPTO.spool.p/US0993966/runat.29122004.130829.9394/app.query.fasta_1.647
-DB=EST -OPMT=fastap -SUFFIX=olig.rst -MINMATCH=0.1 -LOOPT=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTPMT=ptco
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0993966@CGN.1.1.6425@runat.29122004.130829.9394 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
EST :
1: gb_est1:*
2: gb_est2:*
3: gb_hlc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	460	97.9	2142	3	AF289584 Homo sapi
2	295	62.8	865	9	AY412098 Homo sapi
3	270	57.4	931	5	BO653673 AGENCOURT
4	248	52.8	921	5	BO643656 AGENCOURT
5	235	50.0	895	5	BO644360 AGENCOURT
6	230	48.9	996	5	BO652087 AGENCOURT
7	227	48.3	990	5	BO064678 AGENCOURT
8	212	45.1	1008	5	BO646371 AGENCOURT
9	194	41.3	627	4	BT67278 603057995

10	187	39.8	928	5	BO644956 AGENCOURT
11	187	39.8	966	5	BO649813 AGENCOURT
12	181	38.5	953	5	BO846054 AGENCOURT
13	178	37.9	899	5	BO645507 AGENCOURT
14	169	36.0	564	4	BM711145 UI-E-DX1-
15	158	33.6	622	6	CB215756 NISC PR09
16	158	33.6	825	6	CB961961 AGENCOURT
17	151	32.1	1027	4	BG104777 602311726
18	150	31.9	751	4	BG820139 602782356
19	145	30.9	634	7	CN369119 170005322
20	142	30.2	429	7	CN369118 170005322
21	142	30.2	481	4	BI047069 MR3-FN020
22	136	28.9	826	6	CB961668 AGENCOURT
23	130	27.7	1059	5	BQ070932 AGENCOURT
24	104	22.1	777	4	BG544261 602571809
25	101	21.5	885	9	AY412099 Pan trogl
26	96	20.4	654	9	AG112447 Pan trogl
27	96	20.4	931	5	BH856404 AGENCOURT
28	65	13.8	395	1	AI167910 OK29D05.X
29	55	11.7	565	2	BF920913 MR2-NT013
30	49	10.4	465	7	W78547 me79606.X1
31	49	10.4	533	7	CN664615 A0815D03-
32	49	10.4	540	5	BX513216 BX513216
33	49	10.4	674	6	CB056689 NISC J119
34	49	10.4	677	7	CF731367 UI-M-G20-
35	49	10.4	684	7	CK450648 903739 WA
36	49	10.4	724	6	CA328380 UI-M-FY0-
37	49	10.4	888	9	AY412100 Mus muscu
38	49	10.4	913	5	BO895164 AGENCOURT
39	49	10.4	915	6	CB182455 AGENCOURT
40	49	10.4	934	2	BF167269 601775743
41	49	10.4	946	6	BY712252 BY712252
42	49	10.4	1466	3	AK013241 Mus muscu
43	49	10.4	2471	3	AK076164 Mus muscu
44	46	9.8	555	5	BP446790 BP446790
45	45	9.6	434	7	W66677 me19605.X1

ALIGNMENTS

RESULT 1	AF289584	2142 bp	mRNA	linear	HTC 01-JAN-2002
LOCUS	AF289584				
DEFINITION	Homo sapiens clone pp7246 unknown mRNA.				
ACCESSION	AF289584				
VERSION	AF289584.1	GI:18027371			
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Man,D.F. and Gu,J.R.				
TITLE	Novel human cDNA clones with function of inhibiting cancer cell growth				
JOURNAL	Unpublished				
REFFERENCE	2 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Man,D.F. and Gu,J.R.				
AUTHORS	Direct Submission				
TITLE	Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/In 2200 Xie-Tu Road, Shanghai 200032, P. R. China				
JOURNAL	Location/Qualifiers				
FEATURES	1..2142				
source	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="pp7246"				
	277..1689				
	/codon_start=1				
	/product="unknown"				

/protein_id="AAU55768.1"
/db_xref="GI:18027372"
/translation="MKTLSKSPVDVCKRRESPEGDSFVAAAMARKIEIEMIGRORP
GVSQPRQLAGTIGRSIRELVADVLRLTSEEBDFLEVALPEKTDQASGDSE
KMEIRVSEPCPGSKQKLFELQCDVMEEDSRQEWFTLYDPNNKQKVRBDTSL
HTTYEVVDSVNHSPITSKMLRVCLTVAIPGSGSKRVLVNOALQGRAPRAEKPTE
DLSEMEKORAPLRFQGDSELEQSCYHNCVDENIERNNYLDLALENTSGSPDS
PSVAOKSELPRPTSNPRTSRSHPEALHIDPRKQGVDPASPHRLDTPIKVSELROR
LRQTQDSSKIFVSPKAOGSKSVGVGHARAKARNKPRIGRAIPAVSPAHIAASPLALP
SLAVLGHKKHKKHAKESQCGCGRLQAPLASGGVPLGKREHLRELPALVYESQAGPQV
RHEHHHHEHHHHYHFQY"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	2142
Score:	460.00	Matches:	460
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.87%	Indels:	0
DB:	3	Gaps:	0

US-09-993-966-7 (1-470) x AF289584 (1-2142)

Qy 11 ValCyselysaArgArgGluSerProGluGlyAspSerPheAlaValSerAlaAlaTrpAla 30
Db 307 GTGTGCAAGCGCAGAGGAGCCCGGAAGGTGACAGCTTCGCGGAGCGCTGGGCT 366
Qy 31 ArgLysGlyIleGluGluTrpIleGlyValArgGlnArgCysProGlyGlyValSerGlyPro 50
Db 367 CGGAAGGGCATCGAGGAGTGATCGGAGACAGCGCTCCCGGGCGGTCTCCGGAGCC 426
Qy 51 ArgGlnLeuArgLeuAlaGlyThrIleGlyValArgSerThrArgGluLeuValGlyAspVal 70
Db 427 CGACAGCTGGCGTTGGGGGACCATAGCCGGAAGCACCCGGAGCTGTGGGGGAGCGTG 486
Qy 71 LeuArgSerThrLeuSerGluGluGluGluAspPheArgLeuGluValAlaLeuPro 90
Db 487 TTGAGAGACACGCTCACGAGGAAGAGAGAGACGACCTTCGGCTGGAAGTGGCCCTGCT 546
Qy 91 ProGluLysThrArgGlyLeuGlySerGlyAspGluLysMetGluArgValSerGlu 110
Db 547 CTGGAAGAAGCTGACGGGCTGGGACGAGATGAGAAAGAGAGAGAGAGAGAGAGAA 606
Qy 111 ProCysProGlySerLysGlnLeuLysPheGluGluGluGluCysAspValSerMet 130
Db 607 CCTCGCCAGGCTCCAGAGACACTGAAAGTTTAAAGGCTCCAGTGGACGTTCATG 666
Qy 131 GlnGluAspSerArgGlnGluTrpThrPheThrLeuTyraAspPheAspAsnGlyLys 150
Db 667 GAGAGAGACAGCGCGGAGGTGACCTTCACCTGTATGACTTTGACCAACAGCGGAG 726
Qy 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrgluValValAspSerSer 170
Db 727 GTCAACCCGAGGAGCATCACAGCTTGTGACACCACTATAGAGGTGGTGAGACTCTCT 786
Qy 171 ValAsnHisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAsp 190
Db 787 GTCAACCACTCCCAACATCCAGCAAGATGCTGGGGTAAAGCTCACACGGTGGCCCGAT 846
Qy 191 GlySerGlnSerLysArgSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArgPro 210
Db 847 GCGAGCCAGAGCAAGAGAGGAGCTCTTGTCAATAGGCTGACCTGCAAGAGCGCAAGGCC 906
Qy 211 ArgAlaGluThrLysProThrGluAspLeuArgSerTrpGluLysValGlnArgAlaPro 230
Db 907 CGAGCAGAGACCAAGCCCACTGAGAGACTGCGGAGCTGGAGAAAGAGCAGCGAGCCCG 966
Qy 231 LeuArgPheGlnGlyAspSerArgLeuGlnSerGlyCysTyrrHisCysValAsp 250
Db 967 CTCAGGTTCCAGGGTGAAGCGCGCTGAGAGAGTGGCTGCTCAACCAATGGTGGAT 1026
Qy 251 GluAsnIleGluArgArgAsnHisTyrrLeuAspLeuAlaGlyIleGluAsnTyrrThrSer 270

Db 1027 GAGAACATCGAGAGAGAAACCACTACTGATCTCGCCGGAGTAGAAAACCTACAGCTCC 1086
Qy 271 GlnPheGlyProGlySerProSerValAlaGlnLysSerGluLeuProAlaGlyThrSer 290
Db 1087 CAATTGGGGCTGAGCTCCCTCCCTCGGAGCCCAAGAGTCAAGACTGCCCCCGACCTCC 1146
Qy 291 AspProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysPro 310
Db 1147 AATCCCACTGATCTCGCTCCCATAGACCGGAAAGCCATCCACATCCACACCGAAAGCCC 1206
Qy 311 GlnGlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGlu 330
Db 1207 CAAGCGGTGAGCCCGGCTCTCTCCACTTCTTGACACCCCAATCCGCAAGTCTCAGAG 1266
Qy 331 LeuGlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValAspSerProLys 350
Db 1267 CTCAGCAAGAGGCTCCCGGGACCCAGAGAGGAGCAAGCATTTGAGAGTCCCCCAAG 1326
Qy 351 AlAGlnGlyAspSerValGlyValGlyHisValAlaArgGlyValaArgAsnLysProPro 370
Db 1327 GCCCAGGCAAGAGTGTGGGTGTGGCCACGTGGCCAGAGGGCCAAAGAAAGCCCTC 1386
Qy 371 LeuGlyProAlaIleProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeu 390
Db 1387 CTGGAGACCCGACATCCCTCGGCTGTCCCTCCGCGCACCTGGCTGCCAGCCGCGCTTC 1446
Qy 391 LeuProSerLeuAlaProLeuGlyHisLysValLysHisLysHisArgAlaLysGluSerGln 410
Db 1447 CTCCTCTCTTACCCCTCCCGGACCAAGAGCAAGAGCAAGAGCAAGAGAGAGAGAG 1506
Qy 411 GlnGlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyLysProValLeuGlyArgGlu 430
Db 1507 CAGGGCTGGGGGGCTCGAGGACCACTGGCTCAGGTGGCTTCCTGGGGCGGAG 1566
Qy 431 HisLeuArgGluLeuProAlaLeuValValTyrgLysSerGlnAlaGlyLysProValGln 450
Db 1567 CACCTCGGAGAGTGGCCCGCTTGTGTGTATAGAGCCAGGCGCGGAGCGGCTCAG 1626
Qy 451 ArgHisGluHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 1627 AGACATGACACCAACCAACCAATGACATGACATGACATGACATGACATGACATGACAT 1686

RESULT 2
AY412098 885 bp DNA linear GSS 16-DEC-2003
LOCUS
DEFINITION Homo sapiens NKD1 gene, VIRUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY412098
VERSION AY412098.1 GI:39768063
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 885)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.

JOURNAL PUBMED 2 (bases 1 to 885)
REFERENCE Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.

```

FEATURES
  source
    1..885
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      <1..>885
      /gene="NKX1"
      /locus_tag="HCM4444"

ORIGIN

Alignment Scores:
Pred. No.:      2,05e-280      Length:      885
Score:          295.00        Matches:      295
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00%  Mismatches:  0
Query Match:    62.77%        Indels:      0
DB:              Gaps:        0

US-09-993-966-7 (1-470) x AY412098 (1-885)

QY 65 G|L|E|U|V|A|G|Y|A|P|V|A|L|E|U|A|G|A|P|T|H|E|U|S|E|G|I|U|G|I|U|A|S|P|P|H|E|A|G| 84
DB 1 GAGCTCGTGGGGCGACGTTGTGAGACACGCTCAGCGAGAAAGAGAGACGACTTTCCG 60
QY 85 L|E|U|G|I|V|A|L|A|L|E|U|P|P|O|P|R|O|G|I|U|L|Y|E|T|H|A|P|G|I|Y|E|U|G|I|Y|S|E|R|G|I|Y|A|S|P|G|I|U|L|Y|L|Y|S 104
DB 61 CTGGAAGTGGCCCTGCTCTCTGAGAAAGATCGACGCGCTGGCAGCGAGATGAGAAAG 120
QY 105 M|E|C|I|U|A|G|Y|A|S|E|G|I|U|P|P|O|C|Y|E|P|P|O|G|I|Y|S|E|I|Y|S|I|Y|S|G|I|U|L|E|U|Y|P|H|E|G|I|U|L|E|U 124
DB 121 A|T|G|A|G|A|G|A|T|G|A|G|C|G|A|C|C|T|G|C|C|A|G|C|T|C|A|N|A|A|G|C|T|G|A|A|G|T|T|G|A|A|G|A|G|C|T|C 180
QY 125 G|I|N|C|Y|A|S|P|V|A|S|E|T|W|E|T|G|I|U|A|S|P|S|E|A|T|G|I|U|T|P|T|H|P|H|E|T|H|L|E|U|T|Y|A|S|P 144
DB 181 C|A|G|T|G|C|A|C|G|T|C|A|T|G|A|G|A|G|A|C|A|G|C|G|C|A|G|A|G|T|G|A|C|T|T|C|A|C|C|T|G|T|A|T|G|A|C 240
QY 145 P|H|E|A|S|P|A|N|A|N|G|I|Y|L|Y|S|V|A|L|T|H|A|G|I|U|A|S|P|I|L|E|T|H|S|E|I|U|L|E|U|H|I|E|T|H|I|E|T|Y 164
DB 241 T|T|T|G|A|C|A|C|A|C|G|G|A|A|G|T|C|A|C|C|G|A|G|A|C|A|C|A|G|T|T|G|T|G|C|A|C|A|C|C|A|T|C|A|T|C|A|T 300
QY 165 G|I|U|V|A|L|A|S|P|S|E|S|E|S|E|V|A|A|N|H|I|S|E|S|P|R|O|T|H|S|E|S|E|I|Y|S|E|T|L|E|U|A|G|V|A|L|Y|S 184
DB 301 G|A|G|T|G|T|G|A|C|T|C|T|G|T|C|A|C|C|A|C|C|C|C|A|C|A|T|C|C|A|G|A|G|A|T|C|T|G|C|G|G|T|A|A|G 360
QY 185 L|E|U|T|H|V|A|L|A|P|A|O|A|P|G|I|Y|S|E|G|I|N|S|E|I|Y|A|I|S|E|R|V|A|L|E|U|A|A|S|N|G|I|N|A|A|S|P 204
DB 361 C|T|C|A|C|C|T|G|G|C|C|C|G|A|T|G|G|A|G|C|C|A|G|A|G|A|G|A|G|G|C|G|T|T|G|C|A|A|T|C|A|G|C|T|G|A|C 420
QY 205 L|E|U|G|I|N|S|E|I|A|A|P|P|O|A|G|A|G|I|U|T|H|I|Y|E|P|R|O|T|H|G|I|U|A|P|L|E|U|A|G|S|E|T|T|P|G|I 224
DB 421 C|T|G|C|A|G|G|C|C|A|G|G|C|C|C|A|G|C|A|G|A|C|C|A|G|C|C|A|G|C|T|G|G|A|G|C|T|G|G|A|G|C|T|G|G|A|G 480
QY 225 L|Y|E|L|Y|E|G|I|N|A|G|A|P|P|O|L|E|U|A|G|P|H|E|G|I|N|G|I|Y|A|S|P|S|E|A|T|G|I|U|G|I|N|S|E|R|G|I|Y|S 244
DB 481 A|A|A|A|G|A|G|A|G|A|G|C|C|C|G|C|T|C|A|G|G|T|T|C|A|G|G|T|A|C|A|G|C|C|G|C|T|G|A|G|C|A|G|T|C|G|G|C|T|C 540
QY 245 T|Y|T|H|I|E|I|S|C|Y|S|V|A|A|P|G|I|U|A|N|I|E|G|I|U|A|G|I|A|N|H|I|S|Y|T|L|E|U|A|S|P|L|E|U|A|I|G|I 264
DB 541 T|A|C|C|A|C|A|T|T|G|C|G|A|T|G|A|G|A|C|A|T|C|G|A|G|A|G|A|A|C|C|A|C|T|T|A|G|A|T|C|T|G|C|G|G|G 600
QY 265 I|L|E|G|I|U|A|N|T|Y|T|H|S|E|R|G|I|N|P|H|E|G|I|Y|P|R|O|G|I|Y|S|E|P|R|O|S|E|R|V|A|L|A|G|I|N|L|Y|S|E|G|I 284
DB 601 A|T|A|A|A|A|A|C|T|A|C|G|T|C|C|A|A|T|T|T|G|G|C|T|G|G|C|C|C|T|T|C|C|G|G|C|C|A|G|A|A|G|T|C|A|G|A 660
QY 285 L|E|U|P|P|O|A|G|T|H|S|E|A|N|P|P|O|T|H|A|G|S|E|A|T|G|S|E|I|H|I|G|I|U|P|P|O|G|I|U|A|I|L|H|I|S 304
DB 661 C|T|G|C|C|C|C|C|G|C|A|C|T|C|C|A|T|C|C|A|C|T|G|A|T|C|T|G|C|T|C|C|A|T|G|A|C|C|G|A|A|G|C|A|T|C|A|C 720
QY 305 I|L|E|P|P|O|H|I|E|A|I|G|Y|P|P|O|G|I|N|G|I|Y|A|A|P|P|O|A|L|A|S|E|P|H|E|I|S|P|H|E|L|E|U|A|S|P|T|H|P|P|O 324
DB 721 A|T|C|C|A|C|A|C|G|A|A|G|C|C|C|A|A|G|G|G|T|G|A|C|C|G|G|C|T|T|C|C|A|C|T|T|G|C|A|C|C|C|C|A 780
QY 325 I|L|E|A|L|Y|S|V|A|S|E|G|I|U|L|E|U|G|I|N|G|I|U|A|G|I|U|L|Y|T|H|R|G|I|N|A|P|G|I|Y|S|E|I|Y|S|H|I|S 344

```

```

DB 781 A|T|C|C|A|A|G|G|T|C|A|G|C|T|C|A|G|C|A|C|G|G|T|C|C|G|G|G|C|A|C|C|C|A|G|A|C|G|G|A|G|C|A|G|C|A|C 840
QY 345 P|H|E|V|A|I|A|S|P|S|E|P|P|O|Y|S|A|I|A|G|I|N|G|I|Y|L|Y|S|E|S|E|V|A|I|G|Y|A|I|G|I 359
DB 841 T|T|T|G|A|G|T|C|C|C|C|A|G|G|C|C|A|G|G|C|C|A|G|A|G|T|G|G|G|T|G|G|G|C 885

RESULT 3
B0653673
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://Image.lnl.gov
Plate: LCM2481 row: h column: 16
High quality sequence stop: 656.
Location/Qualifiers
1..931
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6284223"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into Scori/XhoI sites using the following 5' adaptor:
GGCAGCAGG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:
Pred. No.:      1.08e-255      Length:      931
Score:          270.00        Matches:      270
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00%  Mismatches:  0
Query Match:    57.45%        Indels:      0
DB:              Gaps:        0

US-09-993-966-7 (1-470) x B0653673 (1-931)

QY 201 A|S|N|G|I|N|A|S|P|L|E|U|G|I|N|S|E|I|A|A|P|P|O|A|G|A|G|I|U|T|H|I|Y|P|P|O|T|H|G|I|U|A|S|P|L|E|U 220
DB 3 A|A|T|C|A|G|G|T|A|C|T|G|C|A|G|G|C|G|A|G|G|C|C|C|G|A|G|A|G|C|C|A|A|G|C|C|A|T|G|A|G|A|C|T|G| 62
QY 221 A|G|S|E|T|T|P|G|I|U|L|Y|S|G|I|N|A|G|A|I|A|P|P|O|L|E|U|A|G|P|H|E|G|I|N|G|I|Y|A|S|P|S|E|A|T|G|L|E|U|I 240
DB 63 C|G|A|G|C|T|G|G|A|A|A|G|A|G|A|G|C|C|C|C|C|T|C|A|G|G|T|T|C|A|G|G|T|G|A|C|A|C|C|C|C|T|G|A|G 122
QY 241 G|I|N|S|E|I|Y|C|Y|S|Y|T|H|I|S|C|Y|S|V|A|A|P|G|I|U|A|N|I|E|G|I|U|A|G|I|U|L|Y|T|H|R|G|I|N|A|P|G|I|Y|S|E|I|Y|S|H|I|S 260

```

```

Db      123  CAGTCTGGCTGCTCCACCACTTGCGTGAAGAAACATCCAGAGAGAAACCACTTCTTA 182
Qy      261  AspleuA1aG1y11eGuAenYrThSerGlnPheG1yProG1ySerProSerVa1a1a 280
Db      183  GATCTCGCCGGGATAGAAAATCAACAGCTCCCAATTTGGGCGCTCCCTTCCGTGGCC 242
Qy      281  G1lySerG1yLeuPProPProArGTrhSerAsnProThzrAsrAsrSerHiG1yPro 300
Db      243  CAGAGTCAGAAACGCCCCCGGCATCTCCATCCATCCATCGATCTCGCTCCCATGACCG 302
Qy      301  G1uA1a1eH1e1leProH1e1sArG1yAsrProG1nG1yVa1aAsrPro1a1eSerPheH1e 320
Db      303  GAAGCATTCCACATCCCAACCGAAAGCCCAAGCGGTGACCCGGCGCTCTTCCACTTC 362
Qy      321  LeuAspThrPro1leAla1ySvalSerG1yLeuG1nG1nArG1yLeuArG1yThG1nAsp 340
Db      363  CTTGACACCCCAATCGCCAAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGGCACCCAGGAC 422
Qy      341  G1ySerThyHisPheValArgSerProLyAlaG1nG1y1ySerValG1yVa1G1yHis 360
Db      423  GGGAGCAGACCTTTGAGGTCCCCCAAGGCCCAAGGAGTGTGGTGGGCAC 482
Qy      361  ValAlaArgG1yAlaArGAsnLyPProProLeuG1yProAla1leProAlaValSerPro 380
Db      483  GTGGCCAGAGGGGCGAAGAAACAGCCCTCTGGGACCCGCCATCCCTCGCGTGTCCCCC 542
Qy      381  SerAlaHisLeuAla1a1eSerProAlaLeuLeuProSerLeuAlaProLeuG1yHisLyS 400
Db      543  TCGGCCCATCGGTGCTGCAGCGCCGCGCTCTCCCTCCCTAGAGCCCTCGGGCACAG 602
Qy      401  LyHisThyHisArGAla1ySvalSerG1nG1nG1ySArG1yLeuG1nAlaProLeu 420
Db      603  AAGCACAAGCACCGAGCCAGAGAGCCAGAGGCGTGGCGGGCGTGCAGGACCACTG 662
Qy      421  AlaSerG1yG1yProVal1leuG1yArG1yHis1eLeuArG1yLeuProAlaLeuVal1a1 440
Db      663  GCCTCAAGTGGCCCTGTCTCGGGGCGGAGCAGCTGGGGAGCTGCCCGCTTGCTGTG 722
Qy      441  TyTG1ySerG1nAlaG1yG1nProVal1G1nArG1ySgluHis1eHis1eHis1eG1uHis 460
Db      723  TATGAGAGCCAGCGCGGCGAGCGCTCGAGACATGACACCAACCATGAACAT 782
Qy      461  HisHisThyThyHis1ePheTyTG1nThr 470
Db      783  CACCACATTAACACCACTTCTACCAACA 812

RESULT 4
B0645656      921 bp      mRNA      linear      EST 15-JUN-2002
LOCUS      AGENCOURT_8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
DEFINITION      5', mRNA sequence.
ACCESSION      B0645656
VERSION      B0645656.1 GI:21769828
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 921)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LCM2484 row: d column: 05

```

```

FEATURES
    source
        High quality sequence start: 16
        High quality sequence stop: 685.
        Location/Qualifiers
            1..921
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6285268"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_100"
                /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into BcclI/XhoI sites using the following 5' adaptor:
                GGCAAGAG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                Library."

ORIGIN
Alignment Scores:
Pred. No.:      5.81e-234      Length:      921
Score:          248.00      Matches:      248
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    52.77%      Indels:      0
DB:             5      Gaps:      0

US-09-993-966-7 (1-470) x B0645656 (1-921)
Qy      202  G1nAlaAspLeuG1nSerAlaArGProArGAlaG1uThrLyPProThzrG1uAspLeuArG 221
Db      25  CAGGCTGACTGTGAGAGCGGAGGCGCCCGGAGAGACCAAGCCCATGAGACCTGGCG 84
Qy      222  SerTrG1uLySvalSerG1nArGAlaProLeuArGPhG1nG1yAsrSerArGLeuG1n 241
Db      85  AGCTGGAGAAAGAAAGCAGCGAGCCCGCTCAGGTTCAGGGGTACAGCGCCCTGGAGCAG 144
Qy      242  SerG1yCySerThyHis1eCySvalAsrG1uAsn1leG1uArGArGAsnHis1eTyTrLeuAsp 261
Db      145  TCTGGCTGTACACCACTTGCATGATGAGAAACATCAGAGGAGAAACCACTTATGAT 204
Qy      262  LeuAlaG1y1leGuAenYrThrSerG1nPhG1yProG1ySerProSerVal1a1G1n 281
Db      205  CTCGCGGATAGAAATCAACAGTCCCAATTTGGGCTGGCTCCCTTCGATGGCCAG 264
Qy      282  LySerG1yLeuPProPProArGTrhSerAsnProThzrAsrAsrSerHiG1yProG1u 301
Db      265  AAGTCAGAACTGGCCCCCGGCACCTCCATCCCATCTGATCTCGCTCCCATGAGCCGGA 324
Qy      302  Ala1leH1e1leProH1e1sArG1yAsrProG1nG1yVa1aAsrPro1a1eSerPheH1ePheLeu 321
Db      325  GCCATCACAATCCACACCCGAAAGCCCAAGGCGGTGAGCCGCGCTCTCCCACTTCTT 384
Qy      322  AspThrPro1leAla1ySvalSerG1yLeuG1nG1nArG1yLeuArG1yThG1nAspG1y 341
Db      385  GACACCCCAATCCGCAAGGTCTCAGAGCTCAGAGAGGCTCCGGGGCACCGAGAGGG 444
Qy      342  SerThyHisPheValArgSerProLyAlaG1nG1yLySerValG1yVa1G1yHisVal 361
Db      445  AGCAAGACATTTGTGAGTCTCCCAAGGCCCAAGGCAAGGTGTGGGTGTGGGCAAGCTG 504
Qy      505  GCGAGAGGGCGCAAGAAACAGCCCTCTGGGAGCCGCGCATCTCCCTGGGTGTCCCTCC 564
Db      382  AlaHisLeuAla1a1eSerProAlaLeuLeuProSerLeuAlaProLeuG1yHis1eLyS 401
Qy      565  GCCCAGCTGGCTGCAGCGCCGCGCTCTCCCTCCCTTACCCCGCTCGGGCACAAAGAG 624
Db      402  HisThyHisArGAla1ySvalSerG1nG1nG1ySArG1yLeuG1nAlaProLeuAla 421

```

Db	635	CACNAGCACCCAGGCCAAGGAAAGCAGAGGGCTTCGCCGGGCTTGCAGGACCACACTGGCC	684
Oy	422	SergIgyIyProValleUgYlArGIUhIsLeuArgGlueUpProAlaLeuValTy	441
Db	685	TCAGGTGGCCCTGTTCTCGTGGGCGGGAGCACTCGCGAGACTGTCCCGCTTGATGTAT	744
Oy	442	GluSerGINalAGlyGlnProVal	449
Db	745	GAGAGCCAGCGCGGCGACCGGTC	768
RESULT 5	B0644360		
LOCUS	B0644360		
DEFINITION	B0644360	895 bp	mRNA linear EST 15-JUL-2002
ACCESSION	AGENCOURT_8351368 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286917		
VERSION	5'', mRNA sequence.		
KEYWORDS	B0644360		
SOURCE	B0644360.1 GI:21768532		
ORGANISM	EST.		
	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/		
AUTHORS	1 (bases 1 to 895)		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rgsabbs-remail.nih.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LNCm2488 row: h column: 22 High quality sequence stop: 622. Location/Qualifiers		
FEATURES	1..895		
source	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6286917"		
	/tissue_type="hepatocellular carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_id="NIH MGC 100"		
	/note="Organ: liver; Vector: pOTB1; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
ORIGIN			
Alignment Scores:			
Pred. No.:	3,95e-221	Length:	895
Score:	235..00	Matches:	235
Percent Similarity:	100..00%	Conservative:	0
Best Local Similarity:	100..00%	Mismatches:	0
Query Match:	50.00%	Indels:	0
DB:	5	Gaps:	0
US-09-963-966-7 (1-470) x B0644360 (1-895)			
Oy	202	GlnAlaEpleuGInSeRAlArGrPrOArGaIaGIUrThLySPrOTHrGlUApLeuArG	221
Db	7	CAGGTCAGCTCGAGAAGCGCAAGGCCCGGAGCAGACCAAGCCACCTCAGAGACTCTGGG	66
Oy	222	serTPgluYlVlyGlnArAlaTroLeuArGrPhGlngIyAspSerArYleuGluGln	241

Db	67	AGCTGGAGGAAGAAAGACAGCAGCCCGCCCTCAGGTTCCAGGGTGAACACCCGCTGGACAG	126
Qy	242	SeSgIyCyErYrThiShiSvYeValAspGluAsnIleGluYrGaYgaAnhiSTYrLeuAsp	261
Db	127	TCTGGCTGCTACCAACATTGGGTGATGAGAACTCAGAGAGAAACCACTTACTTAT	186
Qy	262	LeuAlaGlyIleGluAmnYrThSeriGlnPheIYrProGlySerProSerValAlaGln	281
Db	187	CTCCCGGGAGTGAAGAACTACACCTCCCAATTGGAGCTGACTCCCTTCGGTGGCCAG	246
Qy	282	LysSerGluLeuProProAlaGlyThSerAsnProThrArgSerArgSerhiSgLuProGlu	301
Db	247	AAGTCAGAACTGCCCCCGCCGACTCTCCATTCCTGATCTCGCTCCCATAGACCGGAA	306
Qy	302	AlaIlehiSileProhiSargLYsProGlnGlyValAspProAlaSerPhehiSphLeu	321
Db	307	GCATTCACATCCGACACCGAAACCCGAGGGGTGACCCGGCTCTTCCATTCCTT	366
Qy	322	AspThrProIleAlaLYsValSerGluLeuGlnGlnArgLeuArgLYrThGlnAspGly	341
Db	367	GACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTCCGGGGCACCCAGACGGG	426
Qy	342	SerLYsAspPheValArgSerProLYsAlaGlnGlyLYsSerValGlyValGlyhiSVal	361
Db	427	AGCAGGACCTTTGGAGGTCCCGCAAGGCCAGGGCAAGTGTGGGTGTGGGCACGTG	486
Qy	362	AlaArgGlyAlaArgAsnLYsProProLeuGlyProAlaIleProAlaValSerProSer	381
Db	487	GCCAGAGGGGCAABAAACAAGCCCCCTCTGGGACCCGCCATCCCTGGGGTGTCCCTCC	546
Qy	382	AlaHiSleuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyhiSLeLYs	401
Db	547	GCCACCTCGGTGTCACACCCGCGCCCTCTCCCTCCCTAGACCCCTCCGGGCAACAAAG	606
Qy	402	HiLeYhiSAArgAlaLYsGluSerGlnGlnGlyCYsArgGlyLeuGlnAlaProLeuAla	421
Db	607	CACAGACCCGAGCCAGAGAGAGGACACAGGGCTGCGGGGCTGCAAGCACCACTGGCC	666
Qy	422	SeSgIyGlyProValLeuGlyArgGlnhiSLeuArgGluLeuPro	436
Db	667	TCAAGTGCCCTCTGCTGCGGGCGGAGACACTCGGGAGCTGGCC	711
RESULT 6			
LOCUS	B0652087		
DEFINITION	AGENCOURT_8207495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283110		
ACCESSION	5', mRNA sequence.		
VERSION	B0652087		
KEYWORDS	B0652087.1 GI:21776259		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 966)		
COMMENT	NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabds-remail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L1CM2478 row: j column: 07 High quality sequence stop: 604. Location/Qualifiers 1..966 /organism="Homo sapiens"		

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6283110"
/issue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:

Pred. No.:	3.8e-216	Length:	996
Score:	230.00	Matches:	230
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	48.94%	Indels:	0
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x BQ652087 (1-996)

QY 203 ALaaspleuGInsera1aargProaXgAlaGluThrylsProthrgluAspleuXgSer 222
DB 9 GCTGACCTGCAGAGCGCAGAGCCCGGAGAGAGCAACCACTGAGGAGACTCGGAGC 68
QY 223 TrrglulysylsglnaargalaproleuargPheglngllyaspsaargleuGluSer 242
DB 69 TGGGAGAGAGAGCGAGCGCCGCTCAGGTTCCAGGAGTGCACGCCCTGGAGCAGTCT 128
QY 243 G1yCystrHshisCyValaspGluasn1legluarargaasnh1sTYrLeuAspleu 262
DB 129 GGCGCTACCAACCTTCGTAGATGAGAACATCGAGAGAGAAACCACTACTTAGATCTC 188
QY 263 Alag1yl1egluantYrThrSerGlnPheglYProglYserProserValAlGlnLys 282
DB 189 GCGGAGTAGAAMAACACGCTCCCAATTGGGCGCTGCTCCCTCCGTCGGCCCAAG 248
QY 283 SerGluuProProaYrThrSerAsnProthrgSerArgserHsgluProgluAla 302
DB 249 TCGAATCTGCCCCCGGACCTCCATTCACATCTGCTCCCATAGCCGGAAGCC 308
QY 303 I1ehis1leProh1sargLysProglng1yValaspProalaserPheH1sPheLeuasp 322
DB 309 ATCCACATCCCAACGGAAGCCCAAGGCGTGAACCGGCTCTTCCACTTCTTGAC 368
QY 323 ThrPro1lealalyValserGlnleuGlnlnaargleuarg1yThrglnAspGlySer 342
DB 369 ACCCAATCCGCAAGGTCCTCAGAGCTCCAGCAACGCGCTCCGCGGACCAAGAGGAGC 428
QY 343 LysH1shpheValargserProlyValaglnGlyysSerValGlyValGlyValAla 362
DB 429 AAGGACTTGTGAGGTCCCAAGGCCCAAGGCAAGGTGTGGGTGTGGGCACTGAGCC 488
QY 363 ArgG1y1aargAenLysProProleuGlyProal1a1leProalValaserProserAla 382
DB 489 AGAGGGGCAAGAAACAAGCCCCCTCTGGAGCCGCCATCCCTGGGTGTCCCTCCGCC 548
QY 383 H1sleuAla1aserProalaleuLeuProserleuAlaproleuGlyH1slyeLysH1s 402
DB 549 CACCTGGCTGCGACCGCGCTCTCTCCCTCCCTAGCCCTCCGCGGCAAMAAGACAC 608
QY 403 LysH1shargAlalyserGlnserGlnGlycyValargGlyleuGlnAlaproleuAlaser 422
DB 609 AAGGACGAGCAAGAAGAGCCAGACAGGCTGCGGCGGCTCGAGGACCACTGCGCTCA 668
QY 423 G1yG1yProValleuGlyargGlnH1sleu 432

DB 669 GGTGGCCCTGTCTGGGCGGAGACACTG 698

RESULT 7
BQ064678 990 bp mRNA 1linear EST 02-APR-2002
LOCUS
DEFINITION
AGENCOURT 6853565 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5926427
5', mRNA sequence.
BQ064678
BQ064678.1 GI:19893537
EST.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 990)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strauberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCM2099 row: 1 column: 12
High quality sequence stop: 670.
Location/Qualifiers
1..990
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5926427"
/issue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 99"
/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:	3.48e-213	Length:	990
Score:	227.00	Matches:	227
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	48.30%	Indels:	0
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x BQ064678 (1-990)

QY 201 AsnGln1aaspleuGInsera1aargProaXgAlaGluThrylsProthrgluAspleu 220
DB 3 AATAGGCTACCTGCAGAGCGGAGGCCCGGAGAGAGCAACCACTGAGGACTG 62
QY 221 ArgserTrrglulysylsglnaargalaproleuargPheglngllyaspsaargleuGlu 240
DB 63 CGGAGCTGGAGAGAGAGCGAGCGCCGCTCAGGTTCCAGGAGTGCACCGCGCTGAG 122
QY 241 G1sSerG1yCystrHshisCyValaspGluasn1legluarargaasnh1sTYrLeu 260
DB 123 CAGCTGGCTGCTACCACTTCGTAGATGAGAACATCGAGAGAGAAACCACTACTTA 182
QY 261 AspleuAlag1yl1egluantYrThrSerGlnPheglYProglYserProserValAla 280


```

Db      183 GATTCGCCGGATGAGAACTACACGTCCCAATTTGGGCTCCCTCCGTGACC 242
Qy      281 GILYSSerGluLeuProProlArgThrSerAnProThraSerArgSerHisGluPro 300
Db      243 CAGAGTCAGAACTGCCCCCGGACCTCCCAATCCGATCTCGCTCCCATGAGCCG 302
Qy      301 GUAAlaIleHisIleProHisArgIleProGlnGlyValAspProAlaSerPheHisPhe 320
Db      303 GAAGCCATCCACATCCACACCGAAAGCCCAAGGCGTGAACCGGCTCCCTCCACTTC 362
Qy      321 LeuAspThrProIleAlaIleValSerGluLeuGlnGlnArgLeuArgIleThrGlnAsp 340
Db      363 CTGGACACCCCATCCGCAAGGTCTCAGAGCTCCAGCAAGGCTCCGGGGCAGCCAGAG 422
Qy      341 GlysSerIleHisPheValArgSerProIleValAlaGlnGlyIleSerValGlyValGlyHis 360
Db      423 GGGAGCAGACCTTGTGAGGTCTCCCAAGGCGCAAGGAGCAAGAGTGTGGGTGGGCAC 482
Qy      361 ValAlaArgGlyIleAlaArgAnIleProProLeuGlyProAlaIleProAlaValSerPro 380
Db      483 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGTGGACCCGCAATCCCTGGGTGTCCCC 542
Qy      381 SerAlaHisLeuAlaIleSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIle 400
Db      543 TCGGCCACCTGGCTGCGCACGCCGCTCTCTCCCTCCCTAGGCCCCCTCGGGCAGAG 602
Qy      401 LysHisIleHisIleArgAlaIleGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      603 AACGACAGACGAGGCGAAGAGAGCAGAGGAGGCTGGCGGGCTGACAGGACCACTG 662
Qy      421 AlaSerGlyIleProValLeu 427
Db      663 GCCTCAGGTGGCTGTCTGT 683

```

```

RESULT 8
LOCUS   B0646371              1008 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION
AGENCOURT_8493065 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6300188
5', mRNA sequence.
B0646371
B0646371.1 GI:21770543
EST.
SOURCE  Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1008)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strauberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLCM2514 row: a column: 21
High quality sequence start: 17
High quality sequence stop: 553.
Location/Qualifiers
1..1008
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6300188"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistance)"
/clone_id="NIH_MGC_100"
/notes="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

FEATURES

```

source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6300188"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistance)"
/clone_id="NIH_MGC_100"
/notes="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the Laboratory
of Gerald M. Rubin (University of California, Berkeley)
using Zap-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:

Pred. No.:	Length:	1008
Score:	212.00	225
Percent Similarity:	99.56%	0
Best Local Similarity:	99.56%	0
Query Match:	45.11%	1
DB:	5	0

US-09-993-966-7 (1-470) x B0646371 (1-1008)

```

Qy      202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleProThrGluAspLeuArg 221
Db      25 CAGGCTGACCTGACGAGCGCAAGGCCCGGACAGAGACCAAGCCCACTGAGACTCCGG 84
Qy      222 SerThrGluIleIleGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGln 241
Db      85 AGCTGGAGAAAGAACGACGAGCCCGCTCAGGTTCCAGGAGTGCACGCCCTGGAGCAG 144
Qy      242 SerGlyCysThrHisIleCysValAspGluAsnIleGluArgArgAsnIleIleLeuAsp 261
Db      145 TCTGGCTGCTACCAACCACTTCGTAAGATGAGAACATCGAGAGGAAACCACTACTTAT 204
Qy      262 LeuAlaGlyIleGluLeuIleIleIleIleIleIleIleIleIleIleIleIleIle 281
Db      205 CTCGCCGGAGTAAAGAACTACAGCTCCCAATTTGGGCTGCTCCCTCCGTGGCCAG 264
Qy      282 LysSerGluLeuProProlArgThrSerAnProThraSerArgSerHisGluProGlu 301
Db      265 AAGTCAGAACTGGCCCCCGACCTCCATTCACCTGATCTGCTCCATGAGCCGAA 324
Qy      302 AlaIleHisIleProHisArgIleGlySerGlnGlnGlyValAspProAlaSerPheHisPhe 321
Db      325 GCATTCACATCCCAACCGAAGGCCCAAGGGGTGAGCCGGCTCTCCACTTCTT 384
Qy      322 AspThrProIleAlaIleValSerGluLeuGlnGlnArgLeuArgIleIleIleIle 341
Db      385 GACACCCCAATCCCAAGGCTCAGAGCTCAGCAACGCTCCGGGCGACCCAGGAGCGG 444
Qy      342 SerIleHisPheValArgSerProIleValAlaGlnGlyIleSerValGlyValGlyVal 361
Db      445 AGCAGACACTTGTGAGGTCCCAAGGCCCAAGGCGCAAGGTGTGGGCGACAGTG 504
Qy      362 AlaArgGlyAlaArgAnIleProProLeuGlyProAlaIle-ProAlaValSerPhe 381
Db      505 GCGAGAGGGGCAAGAAACAGCCCTCTGGGACCGCCATCCCTCGGTGTCTCCCTC 564
Qy      381 ValHisIleValAlaIleSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIle 401
Db      565 CGGCCCACTGTGCGCAGCGCCGCGCTCCCTCCCTCCTAGCCCGGCGACAGAA 624
Qy      401 HisIleHisIleArgAlaIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 421
Db      625 GCAAGAGCACTTGTGAGGTCCCAAGGCCCAAGGCGCAAGGTGTGGGCGACAGTG 684
Qy      421 AserGlyIleProVal 426
Db      685 CTCAGGTGGCTGTGT 700

```

RESULT 9

```

LOCUS   B1767278              627 bp      mRNA      linear      EST 25-SEP-2001
DEFINITION
603057995F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207614 5',
mRNA sequence.
B1767278

```

```
VERSION      B1767278.1  GI:15758869
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 627)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabs-remail.nih.gov
              Tissue Procurement: Life Technologies, Inc.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LLM11520 row: n column: 23
              High quality sequence stop: 580.
              Location/Qualifiers
                1..627
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5207614"
                  /lab_host="DH10B"
                  /note="Organ: pooled lung and spleen; Vector: PCMV-SPORTe;
                  Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                  anonymous pool of 24 week female lung, 16 week female
                  spleen, and 20-22 week male spleens. Library is oligo-dT
                  primed and directionally cloned (EcoRV site is destroyed
                  upon cloning). Average insert size 1.4 kb, insert size
                  range 1-3 kb. Library is normalized and enriched for
                  full-length clones and was constructed by C. Gruber
                  (Invitrogen). Research Genetics tracking code 026. Note:
                  this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      9 13e-181      Length:      627
Score:          194.00         Matches:      194
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    41.28%        Indels:        0
DB:              4            Gaps:          0

US-09-993-966-7 (1-470) x B1767278 (1-627)
QY      183 VallysbaurThValAlaProAspGlySerGlnSerValysArGserValIleuValAsnGln 202
DB      2 GTAAAGCTCAACCGTGGCCCCGATGGCAGCAGCAAGAGAGAGCTCTTGTCAATGAG 61
QY      203 AAlaspleuGlnSerAlaArgProAlaGlnGluThrIlyPProThrGluAspleuArGser 222
DB      62 GCTACCTGCAGACCGCAGAGCCCGCAGCAGAGCAAGACCCACTGAGGACCTCGGAGGC 121
QY      223 TrpGluIlyslVsgIlnATGAlaProIleuArpGheGlnGlyAspSerArgLeuGluGlnSer 242
DB      122 TGGAGAGAGAGCAGCGCGCCCGCTCAGGTCAGGGTGACCGCGCTGAGCAGAGCT 181
QY      243 GlyCyTrYrIshIsCySValAspGluAsnIleGluArGArGAsnHsTYrLeuAspleu 262
DB      182 GGTGCTACCAACCAATGGTGTGATGAGACATCGAGAGGAGAAACCACTACTTATGATCTC 241
QY      263 AAlaGlyIleGluArYrThSerGlnPheGlyProGlySerProSerValAlaGlnIly 282
DB      242 GCCCGGATAGAAACCTACACCTCCCAATTGGGCTCGCTCCCTTCGTTGGCCAGAG 301
QY      283 SerGluIeuProProAlaGThrSerAsnProThrArGserArGserHsIleGluProGluAla 302
|||||
```

```
DB      302 TCAGAACTGGCCCCCGCACCTCCAAATCCCATGATCTGCTCCATGAGCGGAGGCC 361
QY      303 ILeHisIleProHisatGlyPProGlnGlyValAspProAlaSerPheHispleuAsp 322
DB      362 ATCCACATCCCAACACCCAAAGGCCCAAGGGGTGAGCCGGCCTCTTCACCTCTTGAC 421
QY      323 ThrProIleAlaIlyValSerGluLeuGlnGlyArGleuArGlyThrGlnAspGlySer 342
DB      422 ACCCAATCCGCAAGGTCTCAGAGCTTCACACACGCTCCGGGCACCCAGAGCGGAGC 481
QY      343 LyHisIlePheValArGserProIlyValGlnGlyIlySerValGlyAlaIlyHisValAla 362
DB      482 AAGACTTTGTGAGGTCTCCCAAGGCCCAAGGAGAGTGAGGTGGGTCAGCTGAGCC 541
QY      363 ArgGlyAlaArGAsnIlyPProProlenGlyProAlaIlePro 376
DB      542 AGAGGGGCAAGAAACAGCCCTCTGGGAGCCGCGCATCCCC 583

RESULT 10
LOCUS      B0644956      928 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8511562 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6297032
5', mRNA sequence.
ACCESSION  B0644956
VERSION    B0644956.1  GI:21769128
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 928)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
              Email: cgabs-remail.nih.gov
              Tissue Procurement: CGAP (Stanford)
              CDNA Library Preparation: Rubin Laboratory
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LUCM2505 row: n column: 09
              High quality sequence start: 17
              High quality sequence stop: 551.
              Location/Qualifiers
                1..928
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:6297032"
                  /tissue_type="hepatocellular carcinoma, cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_1lb="NIH_MGC_100"
                  /note="Organ: liver; Vector: pOT87; Site_1: XhoI; Site_2:
                  EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                  into EcoRI/XhoI sites using the following 5' adaptor:
                  GCGACGAG(G). Size-selected >500bp for average insert size
                  1.8kb. Library constructed by Ling Hong in the laboratory
                  of Gerald M. Rubin (University of California, Berkeley)
                  using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                  II RT (Life Technologies). Note: this is a NIH_MGC
                  Library."

ORIGIN
Alignment Scores:
Pred. No.:      1.08e-173      Length:      928
Score:          187.00         Matches:      236
Percent Similarity: 99.16%    Conservative: 0
Best Local Similarity: 99.16% Mismatches:    1
Query Match:    39.79%        Indels:        2
DB:              5            Gaps:          0
```

US-09-993-966-7 (1-470) x B0644956 (1-928)

```

QY 202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeuArg 221
Db 21 CAGGCTGACCTGGAGAGCGAGGCGCCGAGCAAGACCAAGCCCACTGAGGAGCTCGG 80
QY 222 SerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGln 241
Db 81 AGCTGGAGAGAGAGAGAGCGAGCGCCCGCTCAGGTTCCAGGGGTGACAGCCGCTGGAGCAG 140
QY 242 SerGlyCySerTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeuAsp 261
Db 141 TCTGGCTGCTACCACTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 200
QY 262 LeuAlaGlyIleGluAsnTyrTrpSerGlnPheGlyProGlySerProSerValAlaGln 281
Db 201 CTGCGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTGGCTCCCTTCGTTGGCCAG 260
QY 282 LysSerGluLeuProProArgTrpSerAspProThrArgSerArgSerHisGluProGlu 301
Db 261 AAGTCAGAACTGCGCCCGCCGACCTCCCAATCCCACTGATCTCGCTCCCATGAGCCGAA 320
QY 302 AlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeu 321
Db 321 GCCATCCCATCCCACTCCAGAACGCGCAAGCGGTGAGCCCGGCTCTCTCCACTTCTT 380
QY 322 AspThrProIleAlaLysValSerGluLeuGlnArgLeuArgIleThrGlnAspGly 341
Db 381 GACACCCCAATCCGCAAGAGTCTCAGAGCTCCAGAACGGCTCGGGGACACCGAGAGCGG 440
QY 342 SerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHisVal 361
Db 441 AGCAACACACTTTGTAGGTGTCCTCCCAAGGCCAGGCGCAAGAGTGTGGGTGTGGGCACTG 500
QY 362 AlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSer 381
Db 501 GCCAGAGGGGCAAGAAACAGAGCCCTCTGGAGACCCGCAATCCCTGGGGTGTCCCTCC 560
QY 382 AlaHisLeuAlaAspProAla-LeuLeuProSerLeuAlaProLeuGlyHisLysVal 401
Db 561 GCCACCTGGCTCGAGCGCGGAGN-CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 619
QY 401 HisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAl 421
Db 620 GCACAGAGCCAGGAGCAAGAGAGCCAGAGGCTGCGGGGCTCTGACGACACCACTGGC 679
QY 421 AspArgLysIleProValLeuGlyLysArgGlyHisLeuArgGlyLeuProAlaLeu 438
Db 680 CTCAGGTGGCTGTCTCTGGGGGCGAGCACTGGGGAGCTGCGCCCTCTG 731

RESULT 11
LOCUS B0649813 966 bp mRNA linear EST 15-JUN-2002
DEFINITION AGENCOURT_8302499 NIH_MGC_100 Homo sapiens cDNA IMAGE:6271026
ACCESSION B0649813
VERSION B0649813.1 GI:21773985
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 966)
TITLE NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbsf-rcmail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LNCM2447 row: b column: 19

High quality sequence stop: 583.

Location/Qualifiers

FEATURES

source

```

1. 966
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6271026"
/labs_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_100"
/Note="Organ: liver; Vector: pOT7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAGAGG. Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1,13e-173 Length: 966
Score: 187.00 Matches: 228
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 1
Query Match: 39.79% Indels: 2
DB: Gaps: 0

```

US-09-993-966-7 (1-470) x B0649813 (1-966)

```

QY 202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeuArg 221
Db 5 CAGGCTGACCTGGAGAGCGAGGCGCCGAGCAAGACCAAGCCCACTGAGGAGCTCGG 64
QY 222 SerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGln 241
Db 65 AGCTGGAGAGAGAGAGAGCGAGCGCCCGCTCAGGTTCCAGGGGTGACAGCCGCTGGAGCAG 124
QY 242 SerGlyCySerTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeuAsp 261
Db 125 TCTGGCTGCTACCACTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 184
QY 262 LeuAlaGlyIleGluAsnTyrTrpSerGlnPheGlyProGlySerProSerValAlaGln 281
Db 185 CTGCGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTGGCTCCCTTCGTTGGCCAG 244
QY 282 LysSerGluLeuProProArgTrpSerAspProThrArgSerArgSerHisGluProGlu 301
Db 245 AAGTCAGAACTGCGCCCGCCGACCTCCCAATCCCACTGATCTGGCTCCCATGAGCCGAA 304
QY 302 AlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeu 321
Db 305 GCCATCCCATCCCACTCCAGAACGCGCAAGAGGCTGAGCCCGGCTCTCTCTCTCTCT 364
QY 322 AspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyLysThrGlnAspGly 341
Db 365 GACACCCCAATCCGCAAGAGTCTCAGAGCTCCAGAACGGCTCGGGGACACCGAGAGCGG 424
QY 342 SerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHisVal 361
Db 425 AGCAAGACCTTTGTAGGTGTCCTCCAGAGGCCCAAGAGTGTGGGTGTGGGCACTG 484
QY 362 AlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSer 381
Db 485 GCCAGAGGGGCAAGAAACAGAGCCCTCTGGAGACCCGCAATCCCTGGGGTGTCCCTCC 544
QY 382 AlaHisLeuAlaAspProAla-LeuLeuProSerLeuAlaProLeuGlyHisLysVal 401

```

```

Db      545 GCCACCTGGCTGCAGCCGCGG-CCTCTCCCTCCCTAGCCCTCGGGCAGACAGAA 603
Qy      401 GHHSHVSHIARALALysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuA1 421
Db      604 GCACAAACACCGAGCCAGAGAGAGCCAGAGGCTGCGGGCTGTGAGGACCACTGGC 663
Qy      421 aserGlyGlyProValLeuGlyArgGlu 430
Db      664 CTCAGGTGGCCCTCTCTCGGGCGGGAG 691

RESULT 12
BU846054      953 bp      mRNA      linear      EST 16-OCT-2002
LOCUS      AGENCOURT 10413301 NIH_MGC_109 Homo sapiens cDNA clone
DEFINITION      IMAGE:6580002 5', mRNA sequence.
ACCESSION      BU846054
VERSION      BU846054.1 GI:24030759
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 953)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabs-remail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LNCM2784 row: d column: 18
              High quality sequence stop: 627.
              Location/Qualifiers
                1..953
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6580002"
                /tissue_type="teratocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_1lb="NIH_MGC_109"
                /note="Organ: ovary; Vector: pOT7; Site 1: EcoRI; Site 2:
              XhoI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGACGAG(G). Library constructed by Ling Hong in the
              laboratory of Gerald M. Rubin (University of California,
              Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
              Superscript II RT (Life Technologies). Note: this is a
              NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:      9.41e-168      Length:      953
Score:          181.00      Matches:      181
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    38.51%      Indels:      0
DB:            5      Gaps:      0

US-09-993-966-7 (1-470) x BU846054 (1-953)

Qy      1 MetGlyLeuSerHisSerGlyProAlaAlaValCysLysArgArgGluSerProGluGly 20
Db      224 ATGGGAAACTTCACTCAAGCCGCGCGCTGTGCAAGCCAGAGAGCCCGAAGT 283
Qy      21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyIleGluGluTTPIlleGlyArg 40
Db      284 GACAGCTTCGCGGTGAGGCTGCTGGAGGCGCATCGAGAGTGGATCGGAGA 343

```

```

Qy      41 GlnArgCysProGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db      344 CAGGCTGCCCGGCGGCTGTCTCGGAGCCCGACAGCTGGCTTGGCGGCACCATGGC 403
Qy      61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db      404 CGAAGCACCCGGGAGCTCGTGGCGAGCGTGTGAGAGACAGCTCAGCAGAGAAAGAGAG 463
Qy      81 AspAspPheArgLeuGluValAlaLeuProProGluLeuThrArgGlyLeuGlySerGly 100
Db      464 GACGACTTTCGGGTGGAGAGGCGCTGCTCCCTGAGAGAGCTGACGGCTGGCGAGCGGA 523
Qy      101 AspGlyLysMetGluValSerGluProCysProGlySerGlySerGlyLeuLeu 120
Db      524 GATGAGAGAGAGATGAGAGAGATGAGAGAGACCTGCCAGGCTCCAGAGAGAGCTAGAG 583
Qy      121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
Db      584 TTGGAAGAGCTCCAGTCCAGCTGCACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 643
Qy      141 ThrLeuThrAspPheAspAsnGlnGlyValThrArgGluAspIleThrSerLeuLeu 160
Db      644 ACCCTGATACCTTGACCAACAGGAGAGTCAACCGAGAGACATCACAGCTTCTG 703
Qy      161 HisThrIleTyrGluValAlaAspSerSerValAsnHisSerProThrSerSerGlyMet 180
Db      704 CACACCATCTTAGAGGTGTGAGACTCTCTGTACACCACTCCCAATCATCAGAGATG 763
Qy      181 Leu 181
Db      764 CTG 766

RESULT 13
B0645507      899 bp      mRNA      linear      EST 15-JUL-2002
LOCUS      AGENCOURT 8299367 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271502
DEFINITION      5', mRNA sequence.
ACCESSION      B0645507
VERSION      B0645507.1 GI:21769679
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 899)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabs-remail.nih.gov
              Tissue Procurement: CGAP (Stanford)
              cDNA Library Preparation: Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LNCM2448 row: f column: 15
              High quality sequence stop: 600.
              Location/Qualifiers
                1..899
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6271502"
                /tissue_type="hepatocellular carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_1lb="NIH_MGC_100"
                /note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
              EcoRI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGACGAG(G). Size-selected >500bp for average insert size

```

1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:
 Pred. No.: 8,26-165 Length: 899
 Score: 178.00 Matches: 204
 Percent Similarity: 99.03% Conservative: 0
 Best Local Similarity: 99.03% Mismatches: 0
 Query Match: 37.87% Gaps: 2
 DB: 5 Indels: 0

US-09-993-966-7 (1-470) x B0645507 (1-899)

Qy 202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleuProThrGluAspLeuArg 221
 Db 5 CAGGCTACCTGCGAGCGGAGGCGCCGAGCAGAGCAAGCCACTGAGGACCTGGCG 64
 Qy 222 SerTrrpGluuyluylsGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGluGln 241
 Db 65 AGCTGGGAGAAAGAGCAGCGAGCCCGCTCAGCTTCAGGGTGACAGCCGCTGGACAG 124
 Qy 242 SerGlyCysTrrpHisGlyValAspGluAsnIleGluArgArgAsnHisTrrpLeuAsp 261
 Db 125 TTCTGGCTGCTACCACTTCGTGTGATGAGAACATCGAGAGAGAAACCACTACTTGAT 184
 Qy 262 LeuAlaGlyIleGluAsnTrrpThrSerGlnPheGlyProGlySerProSerValAlaGln 281
 Db 185 CTGCGCGGGATAGAAACTACACGTCCTCAATTTGGCTGGCTCCCTTCGTTGGCCGAG 244
 Qy 282 LysSerGluLeuProProArgTrrpSerAspProThrArgSerArgSerHisGluProGlu 301
 Db 245 AAGTCAGAACTGCGCCCGCCGACCTCCATGCCATGCCATGCCATGCCAGCCGGA 304
 Qy 302 AlaIleHisIleProHisArgGlySerProGlnGlyValAspProAlaSerPheHisPheLeu 321
 Db 305 GCATATCCATCCACCAACCAAGCCCAAGCGCTGAGCCGCGCTCCCTCCACTTCCTT 364
 Qy 322 AspThrProIleAlaValSerGluLeuGlnGlnArgGlyIleThrGlnAspGly 341
 Db 365 GACACCCCAATCGCAAGGCTCTCAGACTCCAGCAAGCGCTCCGGGACACCCAGAGCGG 424
 Qy 342 SerLysHisPheValArgSerProLysAlaGlnGlyLysSer-ValGlyValGlyHisIle 361
 Db 425 AGCAAGACCTTTGTGAGCTCCCAAGCCCAAGCCCAAGAGTGTGGGTGGGCAACT 484
 Qy 361 AlaArgGlyValArgAsnLysProProLeuGlyProAlaIle-ProAlaValSerPro 381
 Db 485 GGCACAGAGGGGAGAAAGAGCCCTCTGGGAGCCGCACTCCCTGGGTGTCCTCCCT 544
 Qy 381 exLysHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIle 401
 Db 545 CCGCCCACTGGCTGGAGCGGCGCTCTCTCCCTCTCTAGCCCTCCGAGCAAGA 604
 Qy 401 YehIshIshIshArg 405
 Db 605 AGCAACAAGCACCGG 618

RESULT 14 564 bp mRNA linear EST 28-FEB-2002
 BM711145
 LOCUS U1-E-DX1-agg-m-24-0-UI-r1 UI-E-DX1 Homo sapiens cDNA clone
 DEFINITION U1-E-DX1-agg-m-24-0-UI 5', mRNA sequence.
 ACCESSION BM711145.1 GI:19024403
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 564)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548

COMMENT

Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.regen.com).
 Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers
 1..564
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="U1-E-DX1-agg-m-24-0-UI"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="U1-E-DX1"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 U1-E-DX1 is a normalized cDNA library containing the
 following tissue(s): fetal eyes. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA,
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT7T3-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AGAATCAAGA. This library was created for the program, Gene
 Discovery in the Visual System, supported by National Eye
 Institute (NEI)."

ORIGIN

Alignment Scores:
 Pred. No.: 4,14e-156 Length: 564
 Score: 169.00 Matches: 169
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.96% Indels: 0
 DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x BM711145 (1-564)

Qy 3 LysLeuHisSerIysProAlaAlaValCysLysArgArgGluSerProGluGlyAspSer 22
 Db 1 AAACCTCATTCCAAAGCGCGCGGTGTGCAAGCGCAGGAGAGCCCGAAGGTGAACGC 60
 Qy 23 PheAlaValSerAlaAlaTrrpAlaArgLysGlyIleGluGluTrrpIleGlyArgGlnArg 42
 Db 61 TTGCGCTGAGCGCTGCTGCTGCTCGAAGGCAATCGAAGAGTGAATCGGAGACACGCC 120
 Qy 43 CysProGlyGlyValSerGlyProArgGlnLeuArgGluAlaGlyThrIleGlyArgSer 62
 Db 121 TGCCTGGGCGGTGTCTCGGAGCCCGCACACTCGGTTGGCGGACCAATAGCGCGAAGC 180
 Qy 63 ThrArgGluLeuValGlyAspValLysArgAspThrIleuSerGluGluGluGluAsp 82

Dh 181 ACCGGAGGCTCGTGGCGACGTTGTGAGAGACACGCTCAGCGAGGAGAGAGACGAC 240
Qy 83 PheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGlyAspGlu 102
Dh 241 TTTCGGCTGGAAGGGCTGCTCTCTGAGAGACTAGCGGGCTGGGACGGAGGAGAG 300
Qy 103 LysLeuMetGluArgValSerGluProCyAProGlySerLysLysGlnLeuLysPheGlu 122
Dh 301 AAGAGATGAG 360
Qy 123 GluLeuGlnCyAAspValSerMetGluGluAspSerArgGlnGluTTPThrPheThrLeu 142
Dh 361 GAGCTTCAGTGGCGACGTTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy 143 TyrAspPheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisThr 162
Dh 421 TATGACTTTTACACAAACGCGAGAGTCAACCGAGAGAGACATCACAGCTTGTCCACACC 480
Qy 163 IleTyrGluValValAspSerSerVal 171
Dh 481 ATCTATAGGTGTGTGACTCTCTGTCTC 507

RESULT 15 CB215756 622 bp mRNA linear EST 06-FEB-2003
LOCUS CB215756 NISC np09e12.y1 NICHD_HS_Ut1 Homo sapiens cDNA clone IMAGE:5937502
DEFINITION 5', mRNA sequence.

ACCESSION CB215756
VERSION CB215756.1 GI:28263948
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
CDNA Library Preparation: CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
info@image.llnl.gov
Plate: L1M13164 row: 1 column: 23
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers

FEATURES
SOURCE 1..622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5937502"
/sex="female"
/tissue_type="normal endometrium, late proliferative
phase, cycle day 13"
/lab_host="DH10B (T1-resistant)"
/clone_id="NICHD_HS_Ut1"
/notes="Organ: uterus; Vector: PCMV-SPORT6.1.cdb (Resgen,
Invitrogen Corporation); Site 1: NCI; Site 2: ECRV;
Cloned unidirectionally from microquantity amounts of mRNA
from normal endometrial tissue (late proliferative phase,
cycle day 13). Average insert size 1.9 kb. Library
constructed by ResGen (Invitrogen Corporation)."

ORIGIN

Alignment Scores: 3.35e-145 Length: 622
Pred. No.: 158.00 Matches: 158
Score: 100.00% Conservative: 0
Percent Similarity: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.62% Indels: 0
DB: 6 Gaps: 0
US-09-993-966-7 (1-470) x CB215756 (1-622)
Qy 313 ValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGln 332
Dh 2 GTGAGCCCGGCTCTCTTCCACTTCTTGAACCCCAATGCCCAAGGCTCAGAGCTCAG 61
Qy 333 GlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysValGln 352
Dh 62 CAACGGCTCGGGGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
Qy 353 GlyLysSerValGlyValGlyHisValAlaArgGlyAlaArgAsnLysProProLeuGly 372
Dh 122 GGCAGAGTGTGGGTGTGGGCGACGTGGCCAGAGGGGCAAGAAAGAGAGAGAGAGAG 181
Qy 373 ProAlaIleProAlaValSerProSerAlaHisLeuAlaIleSerProAlaLeuLeuPro 392
Dh 182 CCGGCCATCCCTCGGCTGCTCCCTCCGCGCCCACTGCTCCAGCCCGGCTCTCTCCC 241
Qy 393 SerLeuAlaProLeuGlyHisLysLysLysHisArgAlaLysGlnSerGlnGlnGly 412
Dh 242 TCCCTAGCCCTCCCTCGGCGACAAAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 301
Qy 413 CysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlnHisLeu 432
Dh 302 TGCCGGGGCTGAGAGCAGCACTGGCTCAGTGGCTGTCTCTCTGGGGGGAGACACTTG 361
Qy 433 ArgGluLeuProAlaLeuValValTyrGluSerGlnAlaGlyGlnProValGlnArgHis 452
Dh 362 CGGAGCTGCGCCCTTGTGGTGTATGAGAGCAGGCGGGGAGCGGCTCCAGAGCAT 421
Qy 453 GluHisHisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHis 470
Dh 422 GAGACACACACACACATGAACATCACACACATTAACACACACTTCTTACAGACA 475

Search completed: December 30, 2004, 13:42:16
Job time : 5653 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 06:25:22, Search time 8161 Seconds
(without alignments)
2723.460 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 470

Sequence: 1 MGKLSKPAVCKRRESPEG.....RHEHHHHHHHHHHFYQT 470

Scoring table:
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-O/cn2 1/USPTO.spool.p/US0993966/runat.29122004.130829.9380/app.query.fasta.1.647
-DB=GenEmbl -OPMT=fastat -SUPRTX=olig.rge -MINMATCH=0.1 -LOOPTC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLGN=200 -THR_SCORE=quality -THR_MIN=1 -MODE=LOCAL -OUTWMT=ptco
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0993966 @CN2 1.1 7406 @runat.29122004.130829.9380 -NCPU=6 -ICPU=3
-NO NMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_ba:.*
2: gb_htg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_dr:.*
10: gb_ro:.*
11: gb_ste:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	1438	9	AF358135 Homo sapi
2	470	100.0	1786	9	AY061883 Homo sapi
3	470	100.0	1788	9	AB062886 Homo sapi
4	470	100.0	2604	9	BC051288 Homo sapi

5	383	81.5	1438	6	AR405917	AR405917 Sequence
6	270	57.4	1053	6	CQ721681	CQ721681 Sequence
7	196	41.7	163319	9	HSAB0303140	AB303140 Homo sapi
8	196	41.7	169773	9	AC007608	AC007608 Homo sapi
9	196	41.7	170791	9	AC007334	AC007334 Homo sapi
10	88	18.7	178790	2	AC145238	AC145238 Pan trogl
11	88	18.7	207486	2	AC145175	AC145175 Pan trogl
12	58	12.3	194869	2	AC145003	AC145003 Papio anu
13	58	12.3	198189	2	AC145043	AC145043 Papio anu
14	49	10.4	1401	6	AX172351	AX172351 Sequence
15	49	10.4	1416	10	AF343352	AF343352 Mus muscu
16	49	10.4	1671	10	AF358134	AF358134 Mus muscu
17	49	10.4	1765	10	BC034838	BC034838 Mus muscu
18	46	9.8	1950	6	AX557538	AX557538 Sequence
19	45	9.6	1731	6	AR405916	AR405916 Sequence
20	45	9.6	2340	6	CQ741468	CQ741468 Sequence
21	44	9.4	181401	2	AC146663	AC146663 Sus scrofa
22	44	9.4	199277	2	AC145445	AC145445 Canis fam
23	37	7.9	208061	2	AC145254	AC145254 Bos tauru
24	32	6.8	232509	2	AC134093	AC134093 Rattus no
25	32	6.8	240965	2	AC126054	AC126054 Mus muscu
26	30	6.4	139145	2	AC146326	AC146326 Felis cat
27	26	5.5	237332	2	AC131846	AC131846 Rattus no
28	24	5.1	212531	2	AC145178	AC145178 Gallus ga
29	22	4.7	2469	6	CQ843478	CQ843478 Sequence
30	22	4.7	2469	9	AK124375	AK124375 Homo sapi
31	21	4.5	181105	5	BX664719	BX664719 Zebrafish
32	17	3.6	1191	6	CQ715890	CQ715890 Sequence
33	17	3.6	1285	6	AR405918	AR405918 Sequence
34	17	3.6	1307	6	AR405919	AR405919 Sequence
35	17	3.6	1309	9	BC004940	BC004940 Homo sapi
36	17	3.6	1682	9	AF358137	AF358137 Homo sapi
37	17	3.6	1798	9	AB062887	AB062887 Homo sapi
38	17	3.6	1885	6	AX780854	AX780854 Sequence
39	17	3.6	1947	10	AF358136	AF358136 Mus muscu
40	17	3.6	1958	9	BC012176	BC012176 Homo sapi
41	17	3.6	1967	10	BC019952	BC019952 Mus muscu
42	17	3.6	58359	2	AC110733	AC110733 Mus muscu
43	17	3.6	99335	2	AC010446	AC010446 Homo sapi
44	17	3.6	162641	2	AC091406	AC091406 Rattus no
45	17	3.6	187289	9	AC116351	AC116351 Homo sapi

ALIGNMENTS

RESULT 1	AF358135	1438 bp	mrna	linear	PRI 29-MAY-2001
LOCUS	AF358135				
DEFINITION	Homo sapiens naked cuticle-1 (NC01) mRNA, complete cds.				
ACCESSION	AF358135				
VERSION	AF358135.1	GI:14211713			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Wharton,K.A., Jr., Zimmermann,G., Rouselet,R. and Scott,M.P.				
TITLE	Vertebrate proteins related to Drosophila Naked Cuticle bind				
JOURNAL	Dev. Biol. 234 (1), 93-106 (2001)				
MEDLINE	2125530				
PUBMED	11356022				
REFERENCE	2 (bases 1 to 1438)				
AUTHORS	Wharton,K.A., Jr., Zimmermann,G. and Scott,M.P.				
TITLE	Submitted (08-MAR-2001) Pathology and Molecular Biology, University				
JOURNAL	of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,				
FEATURES	Dallas, TX 75390-9072, USA				
source	Location/Qualifiers				
	1..1438				
	/organism="Homo sapiens"				
	/mol_type="mrna"				

/db xref="taxon:9606"
 /chromosome="16"
 /map="16q12"
 1..1438
 /gene="NKD1"
 8..1420
 /note="containing EFX domain; similar to Drosophila melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"
 /cds_start=1
 /product="naked cuticle-1"
 /protein_id="AAK57484.1"
 /db_xref="GI:14211714"
 /translation="MGKLSKPAVAVCKRRSPGSDPAVSAAVNRGIERMIGRORCP
 GYSGPROLACTIGSTRBELVDVRLDSEBEDPRLVVALPBEKTDGSGDE
 KMRVSEPCPSKQKLFELQDVSWEDSRQEWFTFLYDFDNKVRREDTSL
 HTLYEVDSSVNSPTSSKMLRYKLVAIPGSSQSKSVLVNQADLOSARPAETKPE
 DLRSWEKQCAPLRFQGDSELEQSGCHVCVDENIERNNHYLDLAGIENTYSGPDS
 PSVAQSELPPTSPNPTRSRSHPEALHIHPRKPGVDPAFPLDPIAVSELDOR
 LRGTQDSKHFRVSPKAGSKSVGVHVARANKPGLGPAIPAVSPAHLAASPLLP
 SLAPLGKHKHKHAKESQGCRCGLQAPLASGVLGRHRLRLPALVVTESQAGPQV
 RHEHHHHHHHHHTYQT"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	1438
Score:	470.00	Matches:	470
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
		Gaps:	0

US-09-993-966-7 (1-470) x AK58135 (1-1438)

QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCylLysArgArgLysSerProGlyGly 20
 DB 8 ATGGGGAAATCTCACTCAAGCCGGCCGCTGTGCAAGCGCAGAGGAGAGCCCGAAGGT 67
 QY 21 ArgSerPheAlaValSerAlaAlaTPrAlaArgArgLysLysGlyGlyGlyGlyGlyGly 40
 DB 68 GACAGCTTCGCGGAGCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 127
 QY 41 GlnArgCysProGlyGlyLysSerGlyProArgGlnLeuArgLysLysLysLysLysLys 60
 DB 128 CAGGCGTCCCGGGCGGCTGCTCGGAGCCCGACAGCTGCGGTGGGGGAGACATAGGC 187
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGlnGlnGlu 80
 DB 188 CGAAGCACCCGGGAGCTCGTGGCGAGCGTGTGAGAGACAGCTCAGCGAGGAAGAGAG 247
 QY 81 AspAspPheArgLeuGlnValAlaLeuProProGlnLysThrArgGlyLysGlySerGly 100
 DB 248 GACACCTTCGCGTGGAGAGTGGCCCTCTCTGAGAGACTGACGGGCTGGGAGCGGA 307
 QY 101 AspGlnLysLysMetCysArgValSerGlyProCysArgProGlySerLysLysLysLys 120
 DB 308 GATGAGAAAGAGATGAGAGAGTACGAAACCTCGCCACAGCTCCAGAGACACTGAAG 367
 QY 121 PheGlnGlnLeuGlnCysAspValSerMetGlnLysAspSerArgGlnGlnGlnGlnGln 140
 DB 368 TTGTGAAAGCTCCAGTGGAGCTGTCCATGAGAGAGACAGCCGGCAGAGTGGACCTTC 427
 QY 141 ThrLeuArgAspPheAspAspAspGlnLysValThrArgLysAspLysLysLysLysLys 160
 DB 428 ACCCTGATAGCTTTGACAAACGAGAGGTCCAGGAGAGGACATCACCAAGCTTCCTG 487
 QY 161 HisThrLysArgValAlaLysSerSerValAsnHisSerProThrSerSerLysMet 180
 DB 488 CACACCATCTATGAGTGGTGACTCTCTCTCAACCACTCCCAACATCCAGAAAGATG 547
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
 DB 548 CTGCGGATAAAGCTCACCGTGGCCCGGATGGCAGCAGAGCAAGAGAGAGTCTCTTGTG 607

QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnThrLysProThrGlnAspLeu 220
 DB 608 AATCAGGCTGACTTCGAGAGCGAGAGCGCCCGAGCAGAGACCAAGCCATGAGAGACTTG 667
 QY 221 ArgSerTrpGlnLysGlnArgAlaProLeuArgPheGlnGlnLysSerArgLeuGln 240
 DB 668 CGAGAGCTGGGAGAAAGAGAGAGCGAGCGCCGCTCAGGTTCAGAGGTGACAGCCGCTGAG 727
 QY 241 GlnSerGlyCysTrpThrHisGlyValAspGlnAsnLysGlnArgAspAsnHisTrpLeu 260
 DB 728 CAGCTGGCTGCTACCACTTCGATGATGAAACATCAAGAGAGAAACCACTACTTA 787
 QY 261 AspLeuAlaGlyLysLysValSerGlnLysSerGlnPheGlyProGlySerProSerValAla 280
 DB 788 GATCTGCCCGGATGAAACCTACAGCTCCCAATTTGGGCTTGCTCCCTCCGTGGCC 847
 QY 281 GlnLysSerGlnLysProProArgThrSerAspProThrArgSerArgSerHisGlnPro 300
 DB 848 CAGAGTCAGAACTGGCCCCCGGACCTCCATCCCATCTGATCTGCTCCATGAGCCG 907
 QY 301 GlnAlaAlaHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 320
 DB 908 GAAAGCATCCACATCCACACCAAGCCCAAGCGCTGAGACCCGAGCTCTCCACTTC 967
 QY 321 LeuAspThrProLysAlaLysValSerGlnLysGlnGlnArgLysArgGlyThrGlnAsp 340
 DB 968 CTTGACACCCCAATCCGACAGGTCTCAGACTCAGACAGCGCTCGGGGACACCCAGGAC 1027
 QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHis 360
 DB 1028 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCCAAGGAGAGTGGGTGGGCGCAC 1087
 QY 361 ValAlaArgGlyLysAlaArgAsnLysProProLeuGlyProAlaLysProAlaValSerPro 380
 DB 1088 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGAGACCCGCAATCTCGTGGTGTCCCC 1147
 QY 381 SerAlaHisLeuAlaLysSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 DB 1148 TCCGCCCACTGGCTGCGAGCCCGGCTCTCTCCCTTACGCCCTCGGGCACAG 1207
 QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgLysLeuGlnAlaProLeu 420
 DB 1208 AAGCACAAGCAGCAGAGCAGAGAGCAGCAGGCTCGCGGAGCTGAGGACCACTG 1267
 QY 421 AlaSerGlyLysProValLeuGlnArgGlnHisLeuArgGlnLysProAlaLeuVal 440
 DB 1268 GCCTCAGGTGGCTCTGTGGGCGGAGCACCCTGCGGAGCTGCGGCTGTGGTGTG 1327
 QY 441 TyrGlnSerGlnAlaGlnProValGlnArgHisGlnHisHisHisHisHisHisHis 460
 DB 1328 TATGAGAGCCAGGCGGGGAGCGGCTCAGAGACATGAGACCAACCAACATGAAAT 1387
 QY 461 HisHisHisTrpHisHisPheTrpGlnThr 470
 DB 1388 CACCACATTACCACTTCTACAGACA 1417

RESULT 2
 AY061883 1786 bp mRNA linear PRI 26-DEC-2001
 LOCUS Homo sapiens naked protein (NKD) mRNA, complete cds.
 DEFINITION
 ACCESSION AY061883
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 1786)
 Yan, D., Wiesmann, M., Rohan, M., Chan, V., Jefferson, A. B., Guo, L.,
 Sakamoto, D., Caothien, R. H., Fuller, J. H., Reinhard, C., Garcia, P. D.,
 Randazzo, F. M., Escobedo, J., Pantl, W. J. and Williams, L. T.
 Elevated expression of axin2 and hnk2 mRNA provides evidence that

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14973-14978 (2001)

PUBMED 11752446

AUTHORS Rohan, M., Chan, V., Yan, D., Fan, L., W. J. and Williams, L. T.

JOURNAL Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA

FEATURES

source
 1. 1786
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1. 1786
 /gene="NKD"
 159. 1571
 /gene="NKD"
 /codon_start=1
 /product="naked protein"
 /protein_id="AAU32374.1"
 /db_xref="GI:17978537"
 /translation="MGKLHSPKPAVCKRRESPEGDSFVSAAMARKGIEEMIGRORCP
 GYSGPRLAAGTIGRSTRVLVDVADRLSEEDPRLVALPEKRDGSGDE
 KMRVSEPCSGSKDLKFEELQDVSEEDSROBFTFYDPDNKGTREDITSL
 HTLYEVVDSSVNSPTSSKMLRVCLVAPDSOSKRSVLVQADLQSKAPRAETKPE
 DLKSEKKQAPLRFQDSRLBQSCVHCVDENIERNNYLDLAGLENTSDFGDS
 PSYAKSELPRTSNPTRSRSEEAHIIHPRKQGVDPASFLDPIAKVELQOR
 LRGTQDSKHFVSPKAGQSGVGVHVARANRPGLPAIPVSPAHIAASPLLP
 SLAPLGKHKHAKESQSGRGIAQLAAGSPVLGRENHRELPAIVVESQAGPQV
 RHHNNHHNNHHNYHNYQT"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	1786
Score:	470.00	Matches:	470
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-993-966-7 (1-470) x AY061883 (1-1786)

QY 1 MetGlyValLeuHisSerLysProAlaAlaValCysAlaArgArgGluSerProGluGly 20
 DB 159 ATGGGGAACTTCACATCCAGCCGCGCGCTGTGCAAGCGAGAGAGCCCGAAGGT 218
 QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgGlyLeuGluGluTTPILLeuArg 40
 DB 219 GACAGCTTCCCGTGAAGCGCTGCTGGAGGCGCATCGAGAGTGGATCGGAGA 278
 QY 41 GlnArgCysProGluGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrLeuGly 60
 DB 279 CACGCGTCCCGCGCGGTCTCTCGGACCCGACAGCTGCGGTTGGCGGACCATAGC 338
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgPheThrLeuSerGluGluGlu 80
 DB 339 CGAAGACCCCGGAGCTCTGGGCGAGCTGTGAGACACAGCTCAGCGAGAGAGAG 398
 QY 81 AspAspPheArgLeuGluValAlaAlaLeuProGluGluThrArgGlyLeuGlySerGly 100
 DB 399 GACGACTTTCGCTGGAGAGTGGCCCTGCTCTGAGAGAGCTGCGGCTGGGCGAGGA 458
 QY 101 AspGluLeuLeuMetGluArgValSerGluProCysArgProGlySerLeuValGluLeu 120
 DB 459 GATGAGAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGAAGAGAGTGAAG 518
 QY 121 PheGluGluLeuGlnCysArgValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
 DB 519 TTGGAAGAGCTTCAGTCCGCTGCTCATGAGAGAGACAGCCGCGAGAGTGAAGCTTC 578
 QY 141 ThrLeuTyrAspPheAspAsnAsnGlyLeuValThrArgGluAspIleThrSerLeuLeu 160

DB 579 ACCCTGATGACTTTGACAAACAGCGCAAGTCAACCCGAGAGACATCACCAGTCTG 638
 QY 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180
 DB 639 CACACCATATGAGGTGGTGGACTCTCTGTCAACCATCCCAATCATCAGCAAGATG 698
 QY 181 LeuArgValIleLeuThrValAlaProAspGlySerGlnSerLeuValLeuVal 200
 DB 699 CTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGCCAGACAGAGAGAGAGAGCTCTTGT 758
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
 DB 759 AATCAGCTAATCTGCGAGCGAGCGAGCCCGGAGACAGACCAAGCCCATGAGAGCTG 818
 QY 221 ArgSerTrpGluValLysGlnArgValProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 DB 819 CGAGCTGGAG 878
 QY 241 GlnSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgAsnHisTyrLeu 260
 DB 879 CAGCTGGCTGCTACCAACCATGCTGATGATGAGAACATCGAGAGAGAGAGAGAGAG 938
 QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 DB 939 GATCTCCCGGATGAAATCAACATCACTCCCAATTTGGCTGGCTCCCTTCCTGGCC 998
 QY 281 GlnLysSerGluLeuProProArgTyrSerAspProThrArgSerArgSerHisGluPro 300
 DB 999 CAGAGTCAGAACTGCCCCCCCCGACCTCAATCCCATCGATCTGCTCCCATGAGCCG 1058
 QY 301 GlnAlaIleHisIleProHisIleGlyLysProGlnGlyValAspProAlaSerPheHisPhe 320
 DB 1059 GAAGCCATCCACATCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
 QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
 DB 1119 CTTGACACCCCAATCCGACAGAGTCTCAGAGCTCAGAGAGCTCGGGGACCCAGAGAC 1178
 QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
 DB 1179 GGGAGCAAGCACTTTGTGAGGTGCCCCCAAGAGCCAGAGAGAGAGAGAGAGAGAGAG 1228
 QY 361 ValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 DB 1239 GTGGCCAGAGAGGAG 1298
 QY 381 SerAlaHisLeuAlaIleSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 DB 1299 TCGGCCACTTGGCTGCGAGCCGAGCCCTCTCTCCCTCTAGCCCCCTCGGACAGAG 1358
 QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 DB 1359 AAGCAGACAGACCGAG 1418
 QY 421 AlaSerGlyLysProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 440
 DB 1419 GCCTCAAGTGGCTGCTCTGCGGCGGAGACCTGCGGAGAGCTGCGCGCTTGGTGGT 1478
 QY 441 TyrGlnSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisGluHis 460
 DB 1479 TATGAG 1538
 QY 461 HisHisHisTyrHisHisPheTyrGlnThr 470
 DB 1539 CACCACCATTAACCACTTCTACAGACA 1568

RESULT 3
 AB062886 1788 bp mRNA linear PRI 20-OCT-2001
 LOCUS Homo sapiens mRNA for Dvl-binding protein NKD1, complete cds.
 DEFINITION
 AB062886
 ACCESSION
 VERSION
 AB062886.1 GI:16303259
 KEYWORDS

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butiria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1
TITLE Molecular cloning, gene structure, and expression analyses of NKD1 and NKD2
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PUBMED 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh,M.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center Research Institute, Genetics and Cell Biology Section; Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan (E-mail:mkatoh@ncc.go.jp, Tel:81-3-3542-2511)
FEATURES
source
1. 1788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/db_xref="GI:16303260"
/protein_id="BAB70500.1"
/translation="MGKLSKPAVACRRSPGDSFVSAAMARKIEEWIGRORCPGGYSPRQILAGTIGRSTRBELVDVRLDTLSEBEDPFLRYALPPEKTDGLSGDEKKMERVSEPCPSKQQLKEBELQCDVMSBEDSRQEWFTLYDFPNNAKVTREDITSLHTTEYVDDSVNHSPTSSKMLRYALIVAPDGSQSKESLVNVDLQGARPRATKPTEDLSWEKKORAPLRFQDGSRLQSGCHNCYDENIERNNYLDLAGIENTYSGPSSPSVAOKSELPRPTSNPTRSRSHPEBAHIHPRKQGVDPASPHLDTPRIAKVSELQORLRGTQDSSKHPSPKAOGSKSVGVGHVBARAKRPGLPRAIPVNSPHNLAAAPLALP SLAPLGHKKHKHAKESQOCGRGLQAPLASGAPVLAGKEHLRELPALVVTESQMGQPVQ RHEHHNHHHHHHYHNPQT"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1788
Score: 470.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-993-966-7 (1-470) x AB062886 (1-1788)
QY 1 MetGlybysleuHisSerIysProAlaAlaValCysIysArgArgIuseProGluGly 20
DB 6 ATGGGAAACTTCACTCCAAAGCCGCGCGCTGTGCAAGGCGAAGGAGCCCGAAAGGT 65
QY 21 AspSerPheAlaValSerAlaAlaTPrAlaArgIysGlyIleGluGluTPrIleGlyArg 40
DB 66 GACAGCTTCGCGCGAAGCGCTGCTGGGCTCGAAGGCGCATCGAGAGTGATCGGAGAGA 125
QY 41 GlnArgCysProGluValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
DB 126 CAGGCGCTGCGCGGCGGTGTCTCGGAGCCCGGACAGCTGCGGTGTGGCGGACCATAGGC 185
QY 61 ArgSerThrArgGlnLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
DB 186 CGAAGCAACCGGAGAGCTCGTGGGCGAGCGTGTGAAGACACGCTCAAGAGAGAGAGAG 245
QY 81 AspAspPheArgGluGluValAlaLeuProProGluIlyThrArgGlyLeuGlySerGly 100
DB 246 GACGACATTTGGCGTGAAGAGTGGCCCTGCTCTGAGAAAGACTGACGGGCTGGGCGAGG 305

QY 101 AspGluIlybysMetGluArgValSerGluProCysProGlySerIysGlnLeuIlys 120
DB 306 GATGAGAGAGAGATGAGAGAGTGAAGGAACTTGGCCAGGCTTCAAGAGAGCTGAAG 365
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTPrPhe 140
DB 366 TTGAAGAGCTCCAGTGCAGCGTGCATGAGAGAGACAGCCGGCGAGTGAACCTTC 425
QY 141 ThrLeuThrAspPheAspAsnAsnGlyValValThrArgGluAspIleThrSerLeu 160
DB 426 ACCCTGTATACCTTGTGCAACCAACGCAAGGTACCCGAGAGACATCACAGCTTCTG 485
QY 161 HisThrIleThrGluValValAspSerSerValAsnHisSerProThrSerSerIysMet 180
DB 486 CACACCATCTATAGAGGTGTGACTCTCTGTCAACCACTCCCAACATCAGAAATG 545
QY 181 LeuArgValIlybysLeuThrValAlaProAspGlySerGlnSerIysArgSerValLeuVal 200
DB 546 CTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGACAGCAAGAGAGCGTCTTGTTC 605
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeu 220
DB 606 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGGAGCAGACCAAGCCCTGAGAGACTG 665
QY 221 ArgSerTrpGluIlybysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
DB 666 CGAAGCTGGAGAGAGAGAGAGCGAGCGCCCTCAGAGTTCAGAGGTGACACCGCTGGAG 725
QY 241 GlnSerGlyCysThrIshisCysValAspGluAsnIleGluArgArgAsnHisTrpLeu 260
DB 726 CAGTGTGGCTGTACCACTTCAGTATGAGAAACATCAAGAGAGAAACCACTACTTA 785
QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
DB 786 GATCTCGCCGGAGTAGAAACTACAGCTCCCAATTGGGCTGGCTCCCTTCGTGGGCC 845
QY 281 GlnIlybysSerGluLeuProProArgGlnSerAspProThrArgSerArgSerHisGluPro 300
DB 846 CAGAGCTCAAACTGCCCCCCCCCAGCTCCCAATCCCACTGATCTGCTCCATGAGCCG 905
QY 301 GlnAlaIleHisIleProHisArgIysProGlnGlyValAspProAlaSerPheHisPhe 320
DB 906 GAAGCCATCCACATCCACACCAAGGCCCCCAAGGCGCTGAGACCGGCTCTCCACTTC 965
QY 321 LeuAspThrProIleAlaIysValSerGlnLeuGlnArgLeuArgGlyThrGlnAsp 340
DB 1086 GTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCCATCTCTGGGTGTCCCCC 1145
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIys 400
DB 1146 TCGGCCACCTGTGTGCACCGCGGCTCTCTCCCTCTCTAGCCCCCTCGGGCACAG 1205
QY 401 IlybHisIlybHisArgAlaIysGlnSerGlnGlnIlyCysArgGlyLeuGlnAlaProLeu 420
DB 1206 AAGCACAAGACCGAGGCAAGAGAGGCGAGGCGTGGGCGCTGGAGGACCACTG 1265
QY 421 AlasSerGlyIlyProValLeuGlyArgGlnHisLeuArgGluLeuProAlaLeuValVal 440
DB 1266 GCCTCAGGTGGCCCTGTCTGGGGCGGAGCACTGTGGAGAGCTGCGGCTGTGTGTG 1325
QY 441 TyrGluSerGlnIleGlyGlnProValGlnArgHisGluHisHisHisHisGluHis 460
DB 1326 TATGAGAGCAGGCGGGGCGAGCCGCTCAGAGACATGAGACCAACCAACCATGAACAT 1385
QY 461 HisHisHisTrpHisHisPheTrpGlnThr 470

Db 705 CACACCATCTATAGAGGCTGGTGAATCCTCTGTGTCAACCACTCCCAACATCCAGCAAGATG 764
Qy 181 LeuArgValIysLeuThrValAlaProAspGlySerGlnSerIlyAspSerValLeuVal 200
Db 765 CTGGGGGTAAAGCTCACCGTGGCCCCGATGGCCAGCAAGAGAGAGAGGCTCTTGTC 824
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgValGlnThrIlyProThrGlnAspLeu 220
Db 825 AATCAGGCTGACCTGCAGAGGCGAAGGCCCGAGACAGAGCCCACTGAGAGACCTG 884
Qy 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlu 240
Db 885 CGGAGCTGGGAGAGAAAGACGAGCCCGCTCAGTTCCAGGGTGAACAGCCGCTGGAG 944
Qy 241 GlnSerGlyCySerThrIlySerGlyValAspGluAsnIleGlnIlyArgAsnIlyThrLeu 260
Db 945 CAGTCTGGCTGTACCACTTCGCTAGATGAGAAACATCGAGAGAGAAACCACTACTTA 1004
Qy 261 AspLeuAlaGlyIleGluAsnThrThrSerGlnPheGlyProGlySerProSerValAla 280
Db 1005 GATCTCCCGGAGTAAAGAACTACAGTCCCAATTTGAGCTGGCTCCCTCCGTGAGCC 1064
Qy 281 GlnIlySerGlyLeuProProArgThrSerAspProThrArgSerArgSerIleGluPro 300
Db 1065 CAGAGCTCAGAACTGCCCGCCCGCACTCCAAATCCCACTGATCTCCCTCCCAAGACCG 1124
Qy 301 GlnAlaIleIleIleProIlyAspGlyProGlnIlyValAspProAlaSerPheIlePhe 320
Db 1125 GAGGCCATCCACATCCCAACCGAAGCCCAAGCCGTGAGCCCGGCTCCCTCCACTTC 1184
Qy 321 LeuAspThrProIleAlaIlySerValSerGlnLeuGlnIlyArgLeuArgIlyThrGlnAsp 340
Db 1185 CTGGACACCCCAATCCGCAAGGCTCAGAGCTCCAGCAACGGCTCCGGAGCAACAGAGAC 1244
Qy 341 GlySerIlySerIlyPheValArgSerProIlySerValGlnIlySerValGlyValIlyIle 360
Db 1245 GGGAGCAAGACCTTTGGAGGTCTCCCAAGGCCCAAGGCAAGATGTGGGTGGGACAC 1304
Qy 361 ValAlaArgGlyAlaArgAsnIlyProProLeuGlyProAlaIleProAlaValAspPro 380
Db 1305 GTGGCCAGAGGGGCAAAACAGCCCTCTGGAGACCCGCAATCCCTGCGGTGTCCCC 1364
Qy 381 SerAlaIleLeuAlaIleSerProAlaLeuProSerLeuAlaProLeuGlyIlyIle 400
Db 1365 TCCGCCCACTGGCTGCGCAGCCCGGCTCTCTCCCTCCCTAGGCCCTCTGGGCAACAG 1424
Qy 401 LysIleIlySerIlySerAlaIlySerGlnGlnIlyCySerArgIlyLeuGlnAlaProLeu 420
Db 1425 AAGCAACAAGCACCGAGCCAGAGGACGAGCGGGCTGGGGCTGGAGGACCACTG 1484
Qy 421 AlaSerGlyIlyProValLeuGlyIlyArgIlyIleLeuArgIlyLeuProAlaLeuVal 440
Db 1485 GCGCTCAGGTGGCTCTGTCTGTGGGCGGAGGACCTGGCGGAGCTGCTGTGTGTG 1544
Qy 441 TyrGlnSerGlnAlaGlyIlyGlnProValGlnArgIleGlnIlySerIlyIleGlnIly 460
Db 1545 TATAGAGCCAGGCGGAGGAGCCGATCCAGAGATGAGCACACACCACTAGAACAT 1604
Qy 461 HisIleIlySerIlyIlySerIlyPheThrGlnThr 470
Db 1605 CACCACCATTAACCACTTCTACAGACA 1634

RESULT 5
AR405917 1438 bp DNA linear PAT 18-DEC-2003
LOCUS AR405917
DEFINITION Sequence 5 from patent US 6630323.
ACCESSION AR405917
VERSION AR405917.1 GI:40154965
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 1438)
AUTHORS Scott, M., Zeng, W., and Wharton, K.
TITLE Naked cuticle genes and their uses
JOURNAL Patent: US 6630323-A 5 07-OCT-2003;
FEATURES
source Location/Qualifiers
1. 1438
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1438
Score: 383.00 Matches: 469
Percent Similarity: 99.58% Conservative: 0
Best Local Similarity: 99.58% Mismatches: 1
Query Match: 81.49% Indels: 2
DB: 6 gaps: 0
US-09-993-966-7 (1-470) x AR405917 (1-1438)
Qy 1 MetGlyIlyLeuIlySerIlySerProAlaIleValCylIlyPheArgArgGlyIlySerProGlnGly 20
Db 8 ATGGGAAACTTCACTCCAAAGCCGCGCGGTGGCAAGCGCAGAGAGCCCGAAAGT 67
Qy 21 AspSerPheAlaValSerAlaIleTTPAlaArgIlyGlyIleGlnIlyTrpIleGlyArg 40
Db 68 GACAGCTTCCCGCTGGAGCGCTGCTGGCTCGGAAGGCAATCGAGAGTGGATCGGAGA 127
Qy 41 GlnArgCySerProGlyIlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db 128 CAGCGTGGCCCGGGCGGTGTCTCGGGACCCCGACAGCTGCGTGGCGGGCACCATTAGGC 187
Qy 61 ArgSerThrArgGlyLeuValGlyIlyAspValLeuArgAspThrLeuSerGlnGlnGlu 80
Db 188 CGAAGCACCCGGAGGCTGTGGGCGAGCTGTGAGACACGCTCAGCGAGAAAGAGAG 247
Qy 81 AspAspPheArgLeuGlnValAlaLeuProProGlnIlySerThrAspGlyLeuGlySerGly 100
Db 248 GACGACTTTGGCTGGAAGTGGCGCTGCTCTGAGAAAGCTGACGGGCTGGGACGGGA 307
Qy 101 AspGlnIlyIlySerMetGlyIlyArgValSerGlnProCySerProGlySerIlyIlyGlnIly 120
Db 308 GATGAGAAAGATGTGAGAGATGAGGAAACCTTGCCCAAGCTCCCAAGAAAGCAAGCTTAAG 367
Qy 121 PheGlnIlyLeuGlnIlyCySerAspValSerMetGlnGlnAspSerArgGlnIlyTrpThrPhe 140
Db 368 TTGAAAGCTCCAGTGCAGCTGTCATGAGAGAGACCGGCGAGGTGACCTTC 427
Qy 141 ThrLeuTyrAspPheAspAsnAsnGlyIlyValThrArgGlnAspIleThrSerLeuLeu 160
Db 428 ACCCTGTATGACTTTGCAACAAACGGAAGTCAACCGAAGGACATCACACGCTTGCTG 487
Qy 161 HisThrIleTyrGlnValValAspSerSerValAsnHisSerProThrSerSerIlyMet 180
Db 488 CACACCATATAGAGGTGGTGAATCTCTGTCAACCACTCCCAACATCAGCAAGATG 547
Qy 181 LeuArgValIlySerLeuThrValAlaProAspGlySerGlnSerIlyAspSerValLeuVal 200
Db 548 CTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCGAGCGCAAGCAAGAGAGAGAGCGTCTTGTC 607
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnThrIlyProThrGlnAspLeu 220
Db 608 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGGAGCAGAGACCAAGCCCACTGAGAGACTG 667
Qy 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlu 240
Db 668 CGGAGCTGGGAGAGAAAGACGAGAGCCCGCTCAGGTTCCAGGGTGAACACCCGCTGGAG 727
Qy 241 GlnSerGlyCySerThrIlySerGlyValAspGluAsnIleGlnIlyArgAsnIlyThrLeu 260
Db 728 CAGTCTGGCTGTACCACTTCGCTAGATGAGAAACATCGAGAGAGAAACCACTACTTA 787
Qy 261 AspLeuAlaGlyIleGluAsnThrThrSerGlnPheGlyProGlySerProSerValAla 280

```

Db      788  GATCTGCGCGGATGAAATCACTCCCATTTGGGCTGCTCCCTCCGTGGCC 847
Qy      281  GlnLysSerGluLeuProProArgThrSer-AsnProThrArgSerArgSerHisGluPr 300
Db      848  CAGAAGTCAGAACTGCGCCCGGACCTT-CAATCCCACTGATCTGCTCCCATGAGCC 906
Qy      300  OGtUAaIleHisIleProHisArgGlySPROGInGlyValAspProAlaSerPheHisPh 320
Db      907  GGAAGCATTCACATCCACACGAAAGCCCAAGGGGTGACCCCGGCTCTCCCTCCACTT 966
Qy      320  eLueAerThrProIleAlaValaSerGluLeuGlnArgGluLeuArgGlyThrGlnAs 340
Db      967  CTTGACACCCCAATGCCCAAGTCTAGAGCTTCAGCAACGGCTCCGGGGCACTCAGGA 1026
Qy      340  pGlySerLysHisPheValaArgSerProLysAlaGlnGlyLysSerValGlyValGly 360
Db      1027  CGGAGCAAGACATTTGTGAGTCCCGCAAGGCCCAAGGCAAGGTGTGGGTGGGCA 1086
Qy      360  eValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPr 380
Db      1087  CGTGGCCAGAGGGGCAAGAAACAGCCCTCTGGGACCGGCATCCCTCGGTGTCC 1146
Qy      380  oSerAlaHisLysValAlaAspSerProAlaLeuLeuProSerLysValaProLeuGlyHis 400
Db      1147  CTCGCCCACTGGCTGCGCCAGCCCGGCTCTCTCCCTCCCTAGCCCCCTCGGACAA 1206
Qy      400  sLysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420
Db      1207  GAAAGCAAGACCAAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1266
Qy      420  uAlaSerGlyGlyProValLeuGlyArgGlyHisLysArgGlyLeuProAlaLeuVala 440
Db      1267  GGGCTAGAGGGCCCTGTCTCTGGGGGAGGACCTGCGGAGCTGCCCTTGAGTGGT 1326
Qy      440  lTyrgLysSerGlnAlaGlyGlnProValGlnArgHisArgLysHisHisHisHisGlu 460
Db      1327  GTATGAGGACCAAGCGCGGAGCGGCTCCAGAGCATGAGACACACACCATGAA 1386
Qy      460  sHisHisHisLysTyrgHisHisPheTyrgLysHis 470
Db      1387  TCACCAACATTCACCAACATTCACCAACATTCACCAACATTCACCAACATTCACCA 1417

RESULT 6
CQ721681 1053 bp DNA linear PAT 03-FEB-2004
LOCUS      CQ721681
DEFINITION Sequence 7615 from Patent WO02068579.
ACCESSION CQ721681
VERSION    CQ721681.1 GI:42282538
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
            Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
            Kites, such as nucleic acid arrays, comprising a majority of
            humenexons or transcripts, for detecting expression and other uses
            thereof
JOURNAL    Patent: WO 02068579-A 7615 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES
            source
            1..1053
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,36e-264 Length: 1053
Score: 270.00 Matches: 270
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

```

Query Match: 57.45% Indels: 0
DB: 6 Gaps: 0
US-09-993-966-7 (1-470) x CQ721681 (1-1053)

Qy      87  ValAlaLeuProProGluLysThrArgGlyLeuGlySerGlyValaAspGlyLysLysMetGlu 106
Db      82  GTGGCCCTGCTCTCTGAGAACTGAGCGGCTGGGCGGAGATGAAAGATGAG 141
Qy      107  ArgValSerGluProCysProGlySerLysLysGlnLeuLysPheGluGluLeuGlnCys 126
Db      142  AGAGTAGGCAACCTGCGCCAGCTCCAGAAAGAGTGAAGTTGAAGAGCTCACTGC 201
Qy      127  AspValSerMetGluLysAspSerArgGlnGluTrpThrPheThrLeuTyraAspPheAsp 146
Db      202  GACGTGTCCATGAGAGGAGGACAGCCCGGACAGAGAGGACCTTCAACCTGTATGACTTTGAC 261
Qy      147  AsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrgLysVal 166
Db      262  AACACGGCAAGGTCAACCGAGAGACATCAAGCTTGCTGACACCATCTATGAGGTG 321
Qy      167  ValAspSerSerValaAsnHisSerProThrSerSerLysMetLeuArgValLysLeuThr 186
Db      322  GTGGACTCCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTGGGGTAAAGCTCAC 381
Qy      187  ValAlaProAspGlySerGlnSerLysArgSerValaLeuValaAsnGlnAlaAspLeuGln 206
Db      382  GTGGCCCCCATGTGGACGCGCAGACCAAGAGAGAGGCTCTGTCAATCAGCTGACCTTCAG 441
Qy      207  SerAlaArgProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGlyLysVala 226
Db      442  AGGCAAGGCGCCGAGAGAGACCAAGCCCATGAGGACCTGGGAGCTGGAGAAAG 501
Qy      227  GlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlnSerGlyCysTyrgHis 246
Db      502  CACGAGAGCCCGCTCAGGTTCCAGGGTGAAGAGCGGCTGGAGCAGTGTGCTGACAC 561
Qy      247  HisCysValaAspGluAsnIleGluArgArgAsnHisTyrgLeuAspLeuAlaGlyLys 266
Db      562  CATTTGGTGATGAGAAACATCGAGAGAGAAACACTACTTAGATTCGCGGGGATGAA 621
Qy      267  AsnTyrgThrSerGlnPheGlyProGlySerProSerValaGlnLysSerGluLeuPro 286
Db      622  AACTACACGTCCCAATTTGGGCTGGCTCCCTTCCTGAGCCCAAGATGAGAACTGCC 681
Qy      287  ProArgThrSerAsnProThrArgSerArgSerHisGluProGluAlaIleHisIlePro 306
Db      682  CCGGCACTCCCAATCCCACTGATCTCGTCCCATGAGCGGAGGCATCCACATCCA 741
Qy      307  HisArgLysProGlnGlyValaAspProAlaSerPheHisPheLeuAspThrProIleAla 326
Db      742  CACCGAAAGCCCAAGCGCGTGGACCCGGCTCTTCCACTTCCCTTGACACCCCAATCGCC 801
Qy      327  LysValSerGluLeuGlnGlnArgLysArgGlyThrGlnAspGlySerLysHisPheVal 346
Db      802  AAGGTCTCAGAGTCTCAGCAAGCGCTCCGGGGACCCAGGACGGAGCAAGCACTTGTG 861
Qy      347  ArgSerProLysAlaGlnGlyLysSerVal 356
Db      862  AGGTCCCGCAAGGCCCAAGGCAAGAGTGTG 891

RESULT 7
HSA303140 16319 bp DNA linear PRI 03-APR-2004
LOCUS      HSA303140
DEFINITION Homo sapiens NOD2 gene for LRR-containing protein, exons 1-11.
ACCESSION AJ303140
VERSION    AJ303140.1 GI:14488148
KEYWORDS   LRR-containing protein; NOD2 gene.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1

```

AUTHORS Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belatche, J., Almer, S., Tyk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sabatou, M. and Thomas, G.

TITLE Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

JOURNAL Nature 411 (6837), 599-603 (2001)

MEDLINE 21279172

PUBMED 11385576

REFERENCE 2 (bases 1 to 163319)

AUTHORS Zouali, H.

TITLE Direct Submission

JOURNAL Submitted (02-JAN-2001) Zouali H., Genetics, Fondation Jean Dausset-CEPH, 27 Rue Juliette Dodu, 75010 Paris France, FRANCE

FEATURES

source

1. 163319

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q12-13"

/clone="hd87b10"

/clone_1fb="CEPH Human BAC library"

98302..131994

/gene="NOD2"

join(98302..98906,106807..106912,109510..111325,115517..115600,115821..115904,118855..118938,121552..121635,122230..122313,124418..124501,128745..128828,130674..131994)

/gene="NOD2"

98302..98906

/gene="NOD2"

/number=1

join(98448..98906,106807..106912,109510..111325,115517..115600,115821..115904,118855..118938,121552..121635,122230..122313,124418..124501,128745..128828,130674..130746)

/gene="NOD2"

/function="putative NF-kB activator"

/note="protein of the CARD family"

/codon_start=1

/product="ILR-containing protein"

/protein_id="CAC42117.1"

/db_xref="GI:1448149"

/db_xref="Swiss-Prot:Q9HC29"

/db_xref="Swiss-Prot:Q9HC29"

/translation="MCSQFAFOAQRSQLVELVSGSLGEPESVLDMLSWELSWEDYEGPHLGQPLSLARLDITVMKGTWACCKLTAAGQAQADSQSPKLHGCMPSHSPADLQSHRAPAIYRIKLSHYENMLDLAWRGFVSQYCDSEIRLPDTPSORARILDLATYKANGLAFLQHVDELPPPLALPLEBATCKYKAKLRTTVSAQSRFLSTYDGAFTLCLEDIYTENNVLEWADVAGSPQSPATLGELEFSTPGHLNDADTVLVGEASGKSTLLQRLHLMAAGQDFQELFVFPFSCROLQCAKPLSVRTLPEHCCMPVQGEIDFQILLHDPRVLTATDGDDEPKRFTDEERHSPDPSPVQTLFNLQGNLLKNA RKVYTSRPAVSAFLRYKIRTERFNLKGFSGOGLIELYRKHNGHGVADRLIRLQETS ALVGTCTLPVPSMWKSCHOELLQEGSGSKTQDWYLLIQLHLAATPDDSAQSLGPSILKRLPRTLHLKRLALMGCMCTYVSAQQLQAQVSPDDISGLFVIRAGVPGSTAPLEFLHITTCFFPAAYLALASADVPAALRHLFNCRPGNSPVARLLPTVCIQASBGDSVVALQKAEPHNQITAAFLAGLSRHWGLABECQTSERALLRQACAM CLARSLRKHFSIIPAPGBAKSVHAPGFIWLSRSLYEQEERLAKARGLNVGHILKLPFCSSVPTCEACALAFVLQHLRRPVALLDYSVGLVGCVKCALYLRNINIDRGICTKLECAHCEQJOKLALFNKLTIDCAHSMATLACRONFLARLGNNTYTAGAQLVLAGLRNTSLQFLGFWNGNVGEGALALDALGDHSLWLSLVGNNTYSVGAQALMLAKNVMLECLTEENHLQDEGCSLAELKKNSSIKLIKLSNNCTTYG ABAALQLAERNDTILEYWRNGTFSLEVDKLGCRDTRLL"

98907..106806

/gene="NOD2"

/number=1

106807..106912

/gene="NOD2"

/number=2

106913..109509

/gene="NOD2"

/number=2

109510..111325

/gene="NOD2"

/number=3

111326..115516

/gene="NOD2"

/number=3

115517..115600

/gene="NOD2"

/number=4

115601..115820

/gene="NOD2"

/number=4

115821..115904

/gene="NOD2"

/number=5

115905..118854

/gene="NOD2"

/number=5

118855..118938

/gene="NOD2"

/number=6

118939..121551

/gene="NOD2"

/number=6

121552..121635

/gene="NOD2"

/number=7

121636..122229

/gene="NOD2"

/number=7

122230..122313

/gene="NOD2"

/number=8

122314..124417

/gene="NOD2"

/number=8

124418..124501

/gene="NOD2"

/number=9

124502..128744

/gene="NOD2"

/number=9

128745..128828

/gene="NOD2"

/number=10

128829..130673

/gene="NOD2"

/number=10

130674..131994

/gene="NOD2"

/number=11

ORIGIN

Alignment Scores:

Pred. No.: 4, 41e-187

Score: 196.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 91.70%

DB: 9

US-09-993-966-7 (1-470) x HSAJ03140 (1-163319)

Qy 275 GlysSerProSerValAlaGlnIlySSerGluLeuProIaIrgThrSerAanProThAArg 294

Db 31966 GGCTCCCTCCCGGCGCCGAGAGTCAAGACTGCCCCCGCACCTCCCAATCCACATCGA 32025

Qy 295 SerArgSerHisGluProGluAlaIleHisIleProHisArgIlySP-roGInGlyValAAsp 314

Db 32026 TCTCGCTCCATGAGCCGGAAGCCATCCACATCCACACCGAAGAGCCGAGGCTGAC 32085

Qy 315 ProIaSerPheHisPheLeuAapThrProIleAlaIyValSerGluLeuGInGlyArg 334

```

Db      32086 CCGGCTCTCTTCCATCTTCTTGTACACCCCAATGCCAAGCTCTAGAGCTTCACCAACGG 32145
Qy      335 LeuArgGlyThrGlnArgGlySerGlyHisPheValArgSerProLysAlaGlnGlyLys 354
Db      32146 CTCGGGGGACCCAGAGCGGGAGACATTTGTGAGGTCTCCCAAGGCCCGGAG 32205
Qy      355 SerValGlyValGlyHisValAlaArgGlyAlaArgAnlyProProLeuGlyProAla 374
Db      32206 AGGTGGGTGTGGGCGCAGTGGCGAGGGGCAAGAACAGCCCTCTGGGACCCGCGC 32265
Qy      375 IleProAlaValSerProSerAlaHisIleuAlaIleSerProAlaLeuLeuProSerLeu 394
Db      32266 ATCCCTGGGAGTGTCCCTCGCCACCTGGCTGGCCAGCCCGGCGCTCTCTCCCTCTA 32325
Qy      395 AlaProLeuGlyHisValGlyHisValArgAlaValArgIleuSerGlnGlnGlyCysArg 414
Db      32326 GCGCCCTCGGGGACAGAAAGCAACAGCACCGAGCCAGAGAGCCAGAGGCTGCGCG 32385
Qy      415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlyHisValLeuArgGly 434
Db      32386 GGCCTGAGGACACCACTGGCTCAGTGGCTGTCTCTGGGGCGGAGCACTGGCGGAG 32445
Qy      435 LeuProAlaLeuValValTyrGlySerGlnAlaGlyGlnProValGlnArgHisGlnHis 454
Db      32446 CTCGCCGCTGTGGTGTGATGAGAGCCAGCGCGGCGAGCGGCTCCAGAGCATGAGCAC 32505
Qy      455 HisHisHisHisGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db      32506 CACCAACACATGACATCACTACCACTTACCACTTCAACAGACA 32553

RESULT 8
AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
LOCUS Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
DEFINITION AC007608
AC007608.7 GI:29124038
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, J., Matibie, M., Bussod, W., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
SUBMITTER Direct Submission
SUBMITTED (20-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
JOURNAL 3 (bases 1 to 169773)
TITLE DOE Joint Genome Institute.
AUTHORS Direct Submission.
JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 4 (bases 1 to 169773)
TITLE DOE Joint Genome Institute.
AUTHORS Direct Submission.
JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 5 (bases 1 to 169773)
TITLE DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
AUTHORS Direct Submission.
JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA

```

```

COMMENT On Mar 19, 2003 this sequence version replaced gi:1829388.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated total Number of Errors is 0.
FEATURES
Source
1. 169773
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-401P9"
ORIGIN
Alignment Scores:
Pred. No.: 4.57e-187 Length: 169773
Score: 196.00 Matches: 196
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.70% Indels: 0
DB: 9 Gaps: 0
US-09-993-966-7 (1-470) x AC007608 (1-169773)
Qy      275 GlySerProSerValAlaGlnLysSerGlyLeuProArgThrSerAspProThrArg 294
Db      78524 GGCCTCCCTTCCTCGGGGCGAGAGTCAAGACTGCCCCCGGACCTCCATCCACACTGA 78465
Qy      295 SerArgSerHisGlyProGlyAlaIleHisIleProHisArgGlyProGlnGlyValAla 314
Db      78464 TTCGCTCCCATGAGCGGAGGACATCCACATCCACACCAAGCCCAAGGCGTGGAC 78405
Qy      315 ProAlaSerPheHisPheLeuAspThrProAlaIleValSerGlyLeuGlnGlnArg 334
Db      78404 CCGGCTCTCTTCCATCTTCTTGAACCCCAATGCGCAAGTCTTCAAGCTTCAGCAACGG 78345
Qy      335 LeuArgGlyThrGlnArgGlySerGlyHisPheValArgSerProLysAlaGlnGlyLys 354
Db      78344 CTCGGGGGACCCAGAGCGGGAGACATTTGTGAGTCTCCCAAGGCCCGGAG 78285
Qy      355 SerValGlyValGlyHisValAlaArgGlyAlaArgAnlyProProLeuGlyProAla 374
Db      78284 AGTGGGTGTGGGCGCAGTGGCGAGGGGCAAGAACAGCCCTCTGGGACCCGCGC 78225
Qy      375 IleProAlaValSerProSerAlaHisIleuAlaIleSerProAlaLeuLeuProSerLeu 394
Db      78224 ATCCCTGGGAGTGTCCCTCGCCACCTGGCTGGCCAGCCCGGCGCTCTCTCCCTA 78165
Qy      395 AlaProLeuGlyHisValGlyHisValArgAlaValArgIleuSerGlnGlnGlyCysArg 414
Db      78164 GCGCCCTCGGGGACAGAAAGCAACAGCACCGAGCCAGAGAGCCAGAGGCTGCGCG 78105
Qy      415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlyHisValLeuArgGly 434
Db      78104 GGCCTGAGGACACCACTGGCTCAGTGGCTGTCTCTGGGGCGGAGCACTGGCGGAG 78045
Qy      435 LeuProAlaLeuValValTyrGlySerGlnAlaGlyGlnProValGlnArgHisGlnHis 454
Db      78044 CTCGCCGCTGTGGTGTGATGAGAGCCAGCGCGGCGAGCGGCTCCAGAGCATGAGCAC 77985
Qy      455 HisHisHisHisGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db      77984 CACCAACACATGACATCACTACCACTTACCACTTCAACAGACA 77937

RESULT 9
AC007334/c 170791 bp DNA linear PRI 13-MAR-2003
LOCUS Homo sapiens chromosome 16 clone RP11-147B17, complete sequence.
DEFINITION AC007334
AC007334.8 GI:28933524
VERSION

```


KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 170791)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 170791)
Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tetum, O.,
Campbell, C., Fawcett, D., McBride, M., Bussod, M., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
REFERENCE 3 (bases 1 to 170791)
DOE Joint Genome Institute.
AUTHORS Direct Submission
JOURNAL Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 170791)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Mar 13, 2003 this sequence version replaced gi:15284265.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.snhg.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1. 170791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-147B17"
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.: 4,596-187 Length: 170791
Score: 196.00 Matches: 196
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.70% Indels: 0
Gaps: 0
US-09-993-966-7 (1-470) x AC007334 (1-170791)
QY 275 GlycerProSerValAlaGlnLysSerGluLeuProAlaGlnThrSerAspProThrArg 294
Db 28069 GGGCTCCCTTCGCGCCGCAAGATCACTCCGCCCGCCCACTCCCAATCCCACTCGA 28010
QY 295 SerArgSerHisGluProGluAlaHisIleProHisArgLysProGlnGlyValAsp 314
Db 28009 TCTGCTCCCAATGAGCCGGAAGCATCCACATCCACACCAAGCCCAAGGGGTGAGC 27950
QY 315 ProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGlnGlyArg 334
Db 27949 CCGGCTCTTCCTCCATCTTGACACACCCCAATCCCAAGGCTCAGAGCTCCACAGCG 27890
QY 335 LeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysValGlnGlyArg 354
Db 27889 CTCGGGGGACCCGAGAGGAGGAGCAAGCACTTTGTGAGGTCCCGCAAGGCCAGGCAAG 27830

QY 355 SerValGlyValAlaGlnHisValAlaArgGlyValAlaArgAsnLysProProLeuGlyProAla 374
Db 27829 AGGTGGGTGTGGCCACGTTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGACCGCC 27770
QY 375 IleProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeuProSerLeu 394
Db 27769 ATCCCTGCGGTGTCCCTCCCTCGGCCACCTTGCTGCGAGCCCGGCTCTCCCTCCTA 27710
QY 395 AlaProLeuGlyHisLysLysLysHisLysLysHisArgAlaLysGluSerGlnGlnGlyArg 414
Db 27709 GCCCGCTCCGGGACAAAGAGCAACAGCACCGGACCAAGAGAGCCGAGGCTGCGG 27650
QY 415 GllLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGly 434
Db 27649 GGCTTGAGGACACCACTGGCTCAGTGAGGCTGTCTTGAGGCGGAGCACTCGGAG 27590
QY 435 LeuProAlaLeuValAlaTyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHis 454
Db 27589 CTCGGCGCTTGGTGTGTATGAGAGCCAGCGCGGCGGTCAGAGACATGACAC 27530
QY 455 HisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 27529 CACCACCAACATGAAATCAACACATTAACACACATTAACACAGACA 27482
RESULT 10
AC145238
LOCUS 178790 bp DNA linear HTG 15-AUG-2003
DEFINITION Pan troglodytes clone CH251-568p19, WORKING DRAFT SEQUENCE, 25
ordered pieces.
AC145238
AC145238.2 GI:33667125
VERSION
ACCESSION
HTG; HTGS PHASE2; HTGS_DRAFT.
KEYWORDS Pan troglodytes (chimpanzee)
SOURCE
ORGANISM
REFERENCE
AUTHORS
Antoniellis, A., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K.,
Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P.,
Hurle, B., Idol, J.R., Karlins, E., Khong, P., Latic, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Marulles, E.H., Masello, C.,
Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.B.,
Prasad, A., Reddix-Bugue, N., Schandler, K., Schueler, M.G., Shah, K.,
Sison, C., Stantirip, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
Vogt, J.L., Wetherby, K.D., Young, A. and Green, E.D.
NISC Comparative Sequencing Initiative
TITLE Unpublished
JOURNAL 2 (bases 1 to 178790)
AUTHORS Green, E.D.
REFERENCE Direct Submission
JOURNAL Submitted (18-JUN-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 178790)
REFERENCE Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
COMMENT On Aug 15, 2003 this sequence version replaced gi:31880081.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@ngri.nih.gov
----- Project Information
Center project name: esk
Center clone name: 568p19

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig have been

established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 171494 bases at least Q40
 Consensus quality: 174120 bases at least Q30
 Consensus quality: 175495 bases at least Q20
 Insert size: 149000; agarose-fp
 Quality coverage: 10.92x in Q20 bases; agarose-fp
 Quality coverage: 9.22x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced

by the finished sequence as soon as it is available and the accession number will be preserved.

1 2604: contig of 2604 bp in length
 2605 2704: gap of unknown length
 2705 7790: contig of 5086 bp in length
 7791 12830: contig of unknown length
 12834 12933: contig of 4943 bp in length
 12934 14664: gap of unknown length
 14665 14764: contig of 1731 bp in length
 14765 32237: contig of 17473 bp in length
 32238 32337: gap of unknown length
 32338 35356: contig of 3019 bp in length
 35357 35456: gap of unknown length
 35457 39282: contig of 3836 bp in length
 39283 39392: gap of unknown length
 39393 41210: contig of 1818 bp in length
 41211 43762: contig of 2452 bp in length
 43763 43862: gap of unknown length
 43863 46766: contig of 2904 bp in length
 46767 46866: gap of unknown length
 46867 53405: contig of 6539 bp in length
 53406 53505: gap of unknown length
 53506 55334: contig of 1829 bp in length
 55335 55434: gap of unknown length
 55435 69095: contig of 13661 bp in length
 69096 69195: gap of unknown length
 69196 78811: contig of 9616 bp in length
 78812 89715: gap of unknown length
 89716 89815: gap of unknown length
 89816 97259: contig of 7444 bp in length
 97260 97359: gap of unknown length
 97360 99238: contig of 1879 bp in length
 99239 99338: gap of unknown length
 99339 101324: contig of 1986 bp in length
 101325 101424: gap of unknown length
 101425 112068: contig of 10644 bp in length
 112069 112168: gap of unknown length
 112169 118679: contig of 6511 bp in length
 118680 118779: gap of unknown length
 118780 120680: contig of 1901 bp in length
 120681 120780: gap of unknown length
 120781 135732: contig of 14952 bp in length

FEATURES

source

* 135733 135832: gap of unknown length
 * 135833 138723: contig of 2891 bp in length
 * 138724 138823: gap of unknown length
 * 138824 163652: contig of 24829 bp in length
 * 163653 163752: gap of unknown length
 * 163753 178790: contig of 15038 bp in length.

Location/Qualifiers

1. 178790
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="CH251-568p19"
 /clone_1b="CH251"
 1. 96615
 /note="clone overlaps with GenBank Accession Number AC145175 clone CH251-32B1 (center project name esj)"
 1. 2604
 /note="assembly_fragment"
 clone_end:Sp6
 vector_side:left"
 2705. 7790
 /note="assembly_fragment"
 7891. 12833
 /note="assembly_fragment"
 12934. 14664
 /note="assembly_fragment"
 14765. 32237
 /note="assembly_fragment"
 32338. 35356
 /note="assembly_fragment"
 35457. 39282
 /note="assembly_fragment"
 39393. 41210
 /note="assembly_fragment"
 41311. 43762
 /note="assembly_fragment"
 43863. 46766
 /note="assembly_fragment"
 46867. 53405
 /note="assembly_fragment"
 53506. 55334
 /note="assembly_fragment"
 55435. 69095
 /note="assembly_fragment"
 69196. 78811
 /note="assembly_fragment"
 78912. 89715
 /note="assembly_fragment"
 89816. 97259
 /note="assembly_fragment"
 97360. 99238
 /note="assembly_fragment"
 99339. 101324
 /note="assembly_fragment"
 101425. 112068
 /note="assembly_fragment"
 112169. 118679
 /note="assembly_fragment"
 118780. 120680
 /note="assembly_fragment"
 120781. 135732
 /note="assembly_fragment"
 135833. 138723
 /note="assembly_fragment"
 138824. 178790
 /note="clone overlaps with GenBank Accession Number AC145174 clone CH251-260A1 (center project name esj)"
 138824. 163652
 /note="assembly_fragment"
 163753. 178790
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"


```

* 152074 160497: contig of 8424 bp in length
* 160498 160597: gap of unknown length
* 160598 170482: contig of 9885 bp in length
* 170483 170583: gap of unknown length
* 170583 207486: contig of 36904 bp in length.
location/Qualifiers
  source
    1. 207486
      /organism="Pan troglodytes"
      /mol_type="genomic DNA"
      /db_xref="taxon:9598"
      /clone="CH251-3281"
      /clone_1fb="CH251"
      1. 430
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:left
      531. 20321
        /note="assembly_fragment"
        20422. 99662
        /note="assembly_fragment"
        96974. 207486
        /note="clone overlaps with GenBank Accession Number
        AC145338 clone CH251-568P19 (center project name esk)"
        99763. 106538
        /note="assembly_fragment"
        106639. 113489
        /note="assembly_fragment"
        113590. 143460
        /note="assembly_fragment"
        143561. 149849
        /note="assembly_fragment"
        149950. 151973
        /note="assembly_fragment"
        152074. 160497
        /note="assembly_fragment"
        160598. 170482
        /note="assembly_fragment"
        170583. 207486
        /note="assembly_fragment"
        missing 17 clone end on 3' end of insert"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 4.19e-77 Length: 207486
Score: 88.00 Matches: 194
Percent Similarity: 97.98% Conservative: 0
Beet Local Similarity: 97.98% Mismatches: 2
Query Match: 18.72% Indels: 4
DB: 2 Gaps: 0

```

US-09-993-966-7 (1-470) x AC145175 (1-207486)

```

QY 275 GYSErProSeRvAlAaGlnLySeRgLuLeuProRArgThrSeRaenProThArg 294
DB 180467 GGGTCCCTTCCCTGCGCCGAGAGTCAAGATCGCCCCCGGCACTTCACCTCGA 180526
QY 295 SeRaSeRHiGlnuProGuaAlaIleHIsIle-ProHIsArguYabRroGlnGlyVala 314
DB 180527 TCTCGCTCCCATGAGCGGAGGACATTCACGT-CCGACACCGAAAGCCCAAGCGCGTGA 180585
QY 314 pProAlaSeRPhHisPheLeuabRhrPrrolleAlaYvalSeRgLuLeuGlnGlnAr 334
DB 180586 CCGGGCTCTTCTCACTTCTTACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACG 180645
QY 334 GLeuASeRGLYThGlnAaPGLYSeRlySHIsPheValaJSeRSeRProLYaAlaGlnGly 354
DB 180646 GCTCCGGGGAGACCAAGACCGGAGCAAGCACTTGTGAGTCCCGCAAGGCCCAAGGGCA 180705
QY 354 eSeRvAlGlyValGlyHIsValAlaArgGlyValaArgAeMlySeRProLeuGlnProAl 374
DB 180706 GAGTGTGGGTGTGGGCACTGCGCCAGAGGGGCAAGAAACAAACCCCTCTGGAGACGGG 180765
QY 374 aIleProAlaValSeRProSeRAla-HIsLeuAlaIaSeRProAlaLeuLeuProSeR 394

```

```

DB 180766 CATTCCTGCGAGTGTCCCTTCCAC-CCACTTGAGTCCAGCCGCGCTCTCCCTCC 180824
QY 394 eulAProLeuGlnHIsLySeRlySHIsPheValaJSeRSeRProLYaAlaGlnGlyCys 414
DB 180825 TAGCCCTCGGGCAGCAAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 180884
QY 414 rGGLYLeuGlnAlaProLeuAlaSeRgLyGlyProValLeuGlyArgGlnHIsLeuArg 434
DB 180885 GGGGGCTGCGAGGACCACTGCGCTCAGGTGCGCTGCTGCTGCGGGGAGACACCTGCGG 180944
QY 434 lLeuProAlaLeuValValTYrGlySeRglnAlaGlnGlnProValGlnArgHIsGln 454
DB 180945 AGCTGCGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 181004
QY 454 lSHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIs 470
DB 181005 ACCACACACACATGAAATCACACATCACACATCACACATCACACATCACACATCAC 181054

RESULT 12
AC145003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Antoniellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Carlsberg,K.,
Chu,G., Coleman,B., Coleman,H., Engle,J., Garg,S., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Marquies,E.H., Mastello,C.,
Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
Prasad,A., Reddix-Dugue,N., Schandler,K., Schneider,M.G., Shah,K.,
Sison,C., Stantrop,S., Thomas,J.W., Thomas,P.J., Tsipour,V.,
Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 194869)
Green,E.D.
Direct Submission
Submitted (30-MAY-2003) NIH Intramural Sequencing Center, 8717
Groveom Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 194869)
Green,E.D.
Direct Submission
Submitted (03-JUL-2003) NIH Intramural Sequencing Center, 8717
Groveom Circle, Gaithersburg, MD 20877, USA
on Jul 3, 2003 this sequence version replaced gi:31193878.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: dtr
Center clone name: 205A15

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average

coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 193209 bases at least Q40
 Consensus quality: 193760 bases at least Q30
 Consensus quality: 193984 bases at least Q20
 Insert size: 180000; agarose-fp
 Quality coverage: 14.34x in Q20 bases; agarose-fp
 Quality coverage: 13.30x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the subcloner.
 * This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 7608: contig of 7608 bp in length
 7609 7708: gap of unknown length
 10151 10151: contig of 2443 bp in length
 10152 10251: gap of unknown length
 10252 44876: contig of 34625 bp in length
 44877 44976: gap of unknown length
 44977 76344: contig of 31368 bp in length
 76345 76444: gap of unknown length
 76445 78547: contig of 2103 bp in length
 78548 78647: gap of unknown length
 78648 145819: contig of 67172 bp in length
 145820 145919: gap of unknown length
 145920 163253: contig of 17334 bp in length
 163254 163353: gap of unknown length
 163354 166118: contig of 2765 bp in length
 166119 166218: gap of unknown length
 166219 194869: contig of 28651 bp in length.

FEATURES

SOURCE

1. 194869
 /organism="Papio anubis"
 /mol_type="genomic DNA"
 /db_xref="taxon:9555"
 /clone="RP41-205A15"
 /clone_lib="RP41"
 1. 31748
 /note="clone overlaps with GenBank Accession Number AC145043 clone RP41-2E13 (center project name dtg)"
 1. 7608
 /note="assembly_fragment
 clone end: 77
 vector side: left"
 7709_10151
 /note="assembly_fragment"
 10252_44876
 /note="assembly_fragment"
 44977_76344
 /note="assembly_fragment"
 76445_78547
 /note="assembly_fragment"
 78648_145819
 /note="assembly_fragment"
 145920_163253
 /note="assembly_fragment"
 163354_166118
 /note="assembly_fragment"
 166219_194869
 /note="assembly_fragment
 clone end: SP6

ORIGIN

vector_side:right"

Alignment Scores: 1.32e-46 Length: 194869
 Pred. No.: 58.00 Matches: 108
 Score: 98.18% Conservative: 0
 Best Local Similarity: 98.18% Mismatches: 1
 Query Match: 12.34% Indels: 2
 DB: 2 Gaps: 0

US-09-993-966-7 (1-470) x AC145003 (1-194869)

Qy 317 SerPheHisPheLeuSerThrProIleAlaValSerGluLeuGlnGlnArg 336

Db 16652 TCCTTCACCTTCCTTACACCCCAATGCCAAGCTCTCAAGCTCCGACAGCGCTCCG 16711

Qy 337 G1YThrGlnApGlySerLyshIsPheValArgSerProLySalGlnGlySerVal 356

Db 16712 GGCACCCAGAGCCGAGCAAGCACTTGTAGGTCCCAAGGCCGAGGCAAGAGCGTG 16771

Qy 357 G1YValGlyHisValAlaArgGlyAlaArgAsnLyseProLeuGlyProAlaIle-Pr 376

Db 16772 GGTGTGGCCACGTGGCCAGAGGGCAAGAAAGCCCTCTGGACCCGCTT-CCC 16830

Qy 376 QAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeuProSerLeuAlaPr 396

Db 16831 TGCAGTCTCCCTCCCTCCGACCTGGCCGAGCCGCGCTCTCCCTCTCTAGCCCC 16890

Qy 396 QLeuGlyHisLyseLyshIsArgAlaLyseGlySerGlnGlnGlyCysArgGlyLe 416

Db 16891 CTTGCGGCAAGAGAGCAAGCAAGCCGAGCAAGAGAGCAGAGGCGTGGCGGCGCT 16950

Qy 416 uGlnAlaProLeuAlaSerGlyPro 425

Db 16951 ACAAGCACCGCTGCGCTCGGTGGCC 16978

RESULT 13

LOCUS AC145043 199189 bp DNA linear HTG 26-JUN-2003

DEFINITION Papio anubis clone RP41-2E13, WORKING DRAFT SEQUENCE, 4 ordered

ACCESSION

AC145043 GI:32261339

KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.

SOURCE Papio anubis (olive baboon)

ORGANISM

Papio anubis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

Cercopithecinae; Papio.

1 (bases 1 to 199189)

REFERENCE

Antoniellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,

Blakesley,R.W., Bonfield,G.G., Brimley,C., Brooks,S., Cariaga,K.,

Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,

Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-D., Hu,P.,

Hurle,B., Idol,J.R., Karins,E., Kwong,P., Latic,P., Lee-Lin,S.-Q.,

Legaspi,R., McDowell,U., Paguirigan,C., Pearson,R., Portnov,M.E.,

Prasad,A., Reddy-Dugue,N., Schandier,K., Scheller,M.G., Shah,K.,

Sison,C., Stantipop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,

Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.

NIHC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 199189)

Green,E.D.

Direct Submission

Submitted (04-JUN-2003) NIH Intramural Sequencing Center, 8717

Grovesmont Circle, Galtersburg, MD 20877, USA

3 (bases 1 to 199189)

Green,E.D.

Direct Submission

Submitted (26-JUN-2003) NIH Intramural Sequencing Center, 8717

Grovesmont Circle, Galtersburg, MD 20877, USA

COMMENT

On Jun 26, 2003 this sequence version replaced gi:31376430.

----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc.zoo@nih.gov
 ----- Project Information
 Center project name: dtg
 Center clone name: 002E13

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 197710 bases at least Q40
 Consensus quality: 197833 bases at least Q30
 Consensus quality: 197878 bases at least Q20
 Insert size: 193000; agarose-fp
 Insert size: 197889; sum-of-contigs
 Quality coverage: 11.84x in Q20 bases; agarose-fp
 Quality coverage: 11.55x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 20663: contig of 20663 bp in length
 * 20664 20763: gap of unknown length
 * 20764 173988: contig of 153225 bp in length
 * 173989 174088: gap of unknown length
 * 174089 176651: contig of 2563 bp in length
 * 176652 176752: gap of unknown length
 * 176752 198189: contig of 21438 bp in length.
 Location/Qualifiers

FEATURES

source

1. 198189
 /organism="Papio anubis"
 /mol_type="genomic DNA"
 /db_xref="taxon:9555"
 /clone="RP41-2E13"
 /clone_lib="RP41"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

1. 20663
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 20764..173988
 /note="assembly_fragment"
 166409..198189
 /note="clone overlaps with GenBank Accession Number
 AC145003 clone RP41-205A15 (center project name der)"
 174089..176651
 /note="assembly_fragment"
 176752..198189
 /note="assembly_fragment"
 missing 77 clone end on 3' end of insert"

Alignment Scores:

Pred. No.: 1.34e-46 Length: 198189
 Score: 58.00 Matches: 108
 Percent Similarity: 98.18% Conservative: 0
 Best Local Similarity: 98.18% Mismatches: 1
 Query Match: 12.34% Indels: 2
 DB: 2 Gaps: 0

US-09-993-966-7 (1-470) x AC145043 (1-198189)

Qy 317 SerPheHisPheLeuAspTrpProIleAlaIysValSerGluLeuGlnGlnArgLysArg 316
 Db 183092 TCCTTCACCTCTCTTGACACCCCAATGCCAAGTCTCAAGCTCCAGCAACGGCTCCG 183151
 Qy 337 GLYThrGlnAspGlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerVal 356
 Db 183152 GGCACCCAGACGGAGCAAGCAACGACTTGTAGAGTCCCCCAAGGCCAGGCAAGAGCGTG 183211
 Qy 357 GLYValGlyHisValAlaArgGlyValArgAspLysProLeuGlyProAlaIle-Pr 376
 Db 183212 GGTGTGGCCACGTCGACAGGGGCAAGAAACAGCCCTCTGGAGACCGCCTT-CCC 183270
 Qy 376 oAlaValSerProSerAlaHisLeuAlaAspProAlaLeuProSerLeuAlaPr 396
 Db 183271 TGCAGTGTCCCTCTCGGCCCACTGGCCGCCAGCCGCCCTCTCCCTCCCTAGCCCC 183330
 Qy 396 oLeuGlyHisLysLysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLe 416
 Db 183331 CTCGGGGCAACAGAACAGCAAGCAGCCAGCCAGAGAGAGCAGAGGGTGGCGGGCCT 183390
 Qy 416 uGlnAlaProLeuAlaSerGlyPro 425
 Db 183391 ACAGGACCGCTGCGCTCGGGTGGCCCC 183418

RESULT 14
 AX172351
 LOCUS AX172351 1401 bp DNA linear PAT 03-JUN-2001
 DEFINITION Sequence 1 from Patent WO0144279.
 ACCESSION AX172351
 VERSION AX172351.1 GI:14597503
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Yan, D. and Williams, L.T.
 TITLE Mammalian dishevelled-associated proteins
 JOURNAL Patent: WO 0144279-A 1 21-JUN-2001;
 Chiron Corporation (US)
 FEATURES
 source 1. 1401
 /organism="Mus musculus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10090"

ORIGIN

Alignment Scores:
 Pred. No.: 2.18e-39 Length: 1401
 Score: 49.00 Matches: 49
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.43% Indels: 0
 DB: 6 Gaps: 0
 US-09-993-966-7 (1-470) x AX172351 (1-1401)
 Qy 131 GluGlnAspSerArgGlnGluTrpThrPheThrLeuValAspPheAspAsnAngLYLys 150
 Db 391 GAGGAGGACAGCGGCAAGAGTGACTTCACTATATGACTTCGACCAATGGCGAAA 450
 Qy 151 ValThrArgGlnAspLierTrpSerLeuLeuHisThrIleTyrGluValValAspSerSer 170

Db 451 GTGACCGGTGAGGACATTACAGCTTGTCGATCATCATATGAGTGGTTGACTCCTCT 510
 QY 171 ValAsnHisSerProThrSerSerIys 179
 Db 511 GTGAACCATTCCTCCCAATCAAGCAAG 537

RESULT 15

AF343352 1416 bp mRNA linear ROD 29-MAR-2001

LOCUS AF343352 Mus musculus Nkd mRNA, complete cds.

DEFINITION AF343352

ACCESSION AF343352.1 GI:13487304

VERSION AF343352.1

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Mus musculus (house mouse)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE 1 (bases 1 to 1416)

JOURNAL Van, D., Wallingford, J. B., Sun, T. Q., Nelson, A. M., Sakanaka, C.,

MEDLINE Reinhard, C., Harland, R. M., Fantl, W. J. and Williams, L. T.

21173639 Cell autonomous regulation of multiple Dishevelled-dependent

11274398 pathways by mammalian Nkd

Proc. Natl. Acad. Sci. U.S.A. 98 (7), 3802-3807 (2001)

REFERENCE 2 (bases 1 to 1416)

AUTHORS Van, D., Wallingford, J. B., Sun, T. Q., Nelson, A. M., Sakanaka, C.,

Reinhard, C., Harland, R. M., Fantl, W. J. and Williams, L. T.

Direct Submission

Submitted (26-JAN-2001) Research, Chiron Corporation, 4560 Horton

Street, Emeryville, CA 94608, USA

FEATURES location/Qualifiers

1..1416

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

1..1416

/note="mNkd; similar to Drosophila naked cuticle"

/codon_start=1

/product="Nkd"

/protein_id="AAK27485.1"

/db_xref="GI:13487305"

/translation="MGKLHSKPAVCKRRSPSGDSPAVSAAMARKGIEEMIGRORCP

GSVSGPROLAGTGVGRGLVGTDRBALGEDDDPFLVAVLPEKIDSLGSGDE

KMERLSEPGQASQKLFRELOCDVSEEDSRQEMFTLYDPNNKQVTRREDITSLL

HTIYEVDSVNSPTSSKTLRVKLTVAPOSQSKSVLFNHTDLOSTRPADTKPAE

ELRGWEKKQAPLRFQGDHLEDPDCAVHCVDENIERRNHYLDLAGIENYTSQFGPS

PSVAQKSELPPRISNPTRRSRSHPEAAHIPPRRPOGVDPGSHLDTPTFAKSELQOR

LRGTODSKHFVSPKAQKMGWGKGARSRKPLVPTTHTVSPSHLATSPALIP

TLAFLGHKKHRAKESQASCRGLQGLAAGSTVMGEQVRELPAVVVYESQAGQAV

ORRHNNHHNNHHHFFQRP"

ORIGIN

Alignment Scores:

Pred. No.: 2,21e-39 Length: 1416
 Score: 49.00 Matches: 49
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.43% Indels: 0
 DB: 10 Gaps: 0

US-09-993-966-7 (1-470) x AF343352 (1-1416)

QY 131 GUGUASPSERARGINGLITPTThPheThIeuTyraSPheAspAAsnGlyLys 150
 Db 391 GAGGAGGACGACCCGCAAGAGTGACTTCACTATATAGACTTCGACAAATGCGCAA 450
 QY 151 ValThrArgGluAspIleThrSerLeuIeuHisThrIleTyrgluValValAspSerSer 170
 Db 451 GTGACCCGTTGAGGACATTACAGCTTGTCATACCATCTATAGAAAGTGCTGACTCCTCT 510
 QY 171 ValAsnHisSerProThrSerSerIys 179
 |||||

Db 511 GTGAACCATTCCTCCCAATCAAGCAAG 537
 Search completed: December 30, 2004, 12:16:23
 Job time : 8674 secs